PARADOXES AND FALLACIES AND THE PROBABILITY AND STATISTICS BEHIND THEM

Byrn Rathgeber

TC 660H
Plan II Honors Program
The University of Texas at Austin

May 4, 2017

Michael Starbird
Department of Mathematics
Supervising Professor

Linda Henderson, Ph.D.
Department of Art and Art History
Second Reader

ABSTRACT

Author: Byrn Rathgeber

Title: Paradoxes and Fallacies and the Probability and Statistics Behind Them

Supervising Professors: Professor Michael Starbird, Dr. Linda Henderson

Probability and statistics form the basis for many of the decisions that we make on a daily basis. However, as often as we weigh the probability of certain events or consequences occurring, we just as often make mistakes in our logic. In this thesis, I examine popular paradoxes and fallacies and seek to explain the mathematical concepts behind them, with the goal of providing a wider audience with examples of common contradictions and mistakes in logic and how to resolve them.

I began by exploring various well-known paradoxes and fallacies in an effort to discover trends in misguided judgment. I selected nine of these and then further examined how they came about and how experts over the decades have aimed to solve them. By researching the solutions, I found that there were overlapping mathematical concepts behind them. I delved into these primary mathematical concepts and discovered that there were three that stood out: basic and conditional probability, expected value theory, and regression to the mean. I then sought to explain these three concepts in a way that would provide readers with the tools to solve the paradoxes in this thesis, as well as similar paradoxes and/or fallacies that they might encounter in the future. Additionally, I discussed real world applications for each of these tools in an effort to demonstrate to the reader how they might incorporate these newly acquired tools into their lives. The thesis concludes by recommending that paradoxes and fallacies be included in a college curriculum, since increased knowledge of them can contribute to better decision-making.

Acknowledgments

I would like to first and foremost thank my Supervising Professors, Professors Starbird and Henderson. I immensely appreciate the continued support you provided throughout this entire process. You came to my rescue and I am forever grateful. Not all heroes wear capes, but the two of you certainly deserve them.

In addition, I would like to thank my family and friends. You saw me through the many ups and downs and loved me through them all. I am forever in debt to many of you for reading pages and pages of paradoxes and fallacies!

Special thanks goes out to Dr. Wettlaufer for her valuable advice and encouragement. I would also like to acknowledge Dr. James Scott for helping to direct me in the preliminary stages of my thesis.

Lastly, I would like to thank Plan II for a wonderful four years. I am better for every quirky and brilliant person I have met and class that I have taken, even Plan II Physics. To my Plan II friends, the late nights have come to an end, at least for now, thanks for the conversations and props to you all!

TABLE OF CONTENTS

SECTION I: INTRODUCTION

"That's All Folks"

Though it has been over seventy years since the witty rabbit crunching a carrot first found his way onto the television screen, the closing lines and Bugs Bunny himself remain iconic.[1] The wisecracking character ran circles around most of his opponents, but was no match for Cecil Turtle. Cecil always managed to outrace, or rather out-trick, Bugs, even if just by a hare.[2]

The fable of the tortoise and the hare has been analyzed and adapted by many, including the famous Greek philosopher, Zeno of Elea. In Zeno's version, the Greek warrior Achilles and a tortoise challenge one another to a race. Suppose Achilles can run 1,000 yards a minute and the tortoise can run 100 yards a minute. Before the race, Achilles is feeling confident and decides to generously give the tortoise a 1,000 yard head start. In a race between a warrior and a tortoise, it seems that the warrior would be the obvious winner. However, after Achilles makes it 1,000 yards in the race, the tortoise is still 100 yards ahead of him, so will Achilles ever catch up?[3]

Zeno would say 'slower when running will never be overtaken by the quicker; for that which is pursuing must first reach the point from which that which is fleeing started, so that the slower must necessarily always be some distance ahead'[4]. In other words, luck is on the side of the tortoise because in order to go a yard, Achilles must first run half of a

---

[1] Katie Nodjimbadem, "What Gives Bugs Bunny His Lasting Power?," Smithsonian.com, July 27, 2015. (The short quote from this source is also included in the conclusion.)
[2] "Bugs Bunny - Tortoise Beats Hare," Cartoonsonnet, accessed 2017.
[3] Edna E. Kramer, *The Nature and Growth of Modern Mathematics* (Greenwich, CT: Fawcett Publications, 1970).
[4] Rachel Thomas, "Mathematical Mysteries: Zeno's Paradoxes," Mathematical Mysteries: Zeno's Paradoxes, December 1, 2000, accessed 2017.

yard, and before that, a quarter of a yard, and before that, an eighth of a yard, and so on to infinity. This series of infinitesimally smaller increments of position is known as a geometric series. The sum of this geometric series, where the multiplier is less than one, will then give the distance Achilles must travel to catch the tortoise. The sum of such a sequence can be found, so the distance Achilles must travel will be finite, and if the distance he must travel is finite, then there is a point at which Achilles could catch up. In addition, if the distance is finite, then the time it will take him to travel it will also be finite. How can this be if there are an infinite number of distances Achilles must travel? This is the paradox. Zeno is thought to have produced 40 of these contradicting and quizzical puzzles, which later came to be known as paradoxes. Zeno's paradoxes, in particular, led to the further questioning and exploration of the concept of motion and whether or not an object is in motion or simply in different static positions over time.

The Achilles and tortoise paradox relies on the fact that distance and therefore time may be infinitely divisible into smaller and smaller pieces. Despite the necessity of this assumption in the paradox, Zeno later proves that distance and time are not infinitely divisible. Additionally, through the arrow paradox, Zeno proves that the universe itself cannot be broken into finite, indivisible elements. The arrow paradox makes the argument that if an arrow is in a place just its own size, then it must be at rest. However, every moment that the arrow is flying it is in a place of its own size. Therefore, by this logic, when an arrow is flying, it must be at rest. This is a contradiction because how can an arrow be flying if it is at rest? Aristotle offered the solution that time is not composed of a set of indivisible instants and that motion and rest do not exist in an instant.[5] Together these paradoxes contributed to the Theory of Relativity and the consideration of light as a particle and as a wave.[6]

Scientific American writer, Martin Gardner, described paradoxes as truths that cut so strongly against the grain of common sense that they are difficult to believe, even after one is confronted with their proofs.[7] In fact, since their beginning, paradoxes have arisen from contradicting truths, which have called for further explanation and investigation into new fields. The questions that they have left unanswered have inspired and forced many to challenge what they know to be true and to seek new truths, which has led to advancements in numerous fields, including those of philosophy, mathematics, physics, finance and more. While they have been valuable throughout history for large discoveries, such as the concept of light as a particle and a wave, they are also relevant to many of the daily decisions that the average person may face, from which investment to make to the interpretation of medical test results. Gaining a better understanding of paradoxes and fallacies and a few of the key mathematical concepts that lead to their solutions would provide even those with no quantitative background better tools for making decisions and, ideally, encourage them to make more successful ones.

In order to find a way to help people make better decisions through the understanding of several mathematical concepts, this thesis will begin by exploring a number of the more popular paradoxes and fallacies. The stories and histories behind each

---

[5] Marc S. Cohen, "Zeno's Paradox of the Arrow," Zeno's Paradox of the Arrow, 2003, accessed 2017.

[6] Ibid.4.

[7] Martin Gardner, *Hexaflexagons and Other Mathematical Diversions: The First Scientific American Book of Puzzles & Games* (Chicago: Univ. of Chicago Press, 1989).

paradox and fallacy will ask the reader to notice why certain situations present conflicting solutions and to note that there may be common errors that occur in reasoning.

After discussing the stories behind nine different paradoxes and fallacies, I will move on to address several key mathematical concepts. These mathematical concepts, including basic and conditional probability, expected value theory, and regression to the mean will supply a person with the background for understanding how to solve and evaluate certain scenarios. These concepts overlap in the paradoxes and fallacies that are discussed in this paper, but are also particularly prevalent in the solutions of other paradoxes and in certain real world circumstances.

The final chapter of this thesis will cover the real world implications of paradoxes and fallacies. A careful description of when contradictions and errors occur in a real-life setting will guide a reader to understand when and how they may be able to utilize the mathematical concepts covered in the previous section. The concepts can be valuable if people find themselves on a game show, in a doctor's office, discussing sports, or buying stocks. Regardless of the situation, picking up a few important mathematical tools should give individuals the ability to solve certain dilemmas and recognize errors in judgment. As a result, these individuals may then be able to make more rational and beneficial decisions.

SECTION II: PARADOXES AND FALLACIES

Basic and Conditional Probability Paradoxes

The Birthday Paradox

Richard von Mises was an Austrian-born American mathematician, engineer, and positivist philosopher known for his contributions to statistics and probability theory. At the beginning of his career, von Mises gave the first German university course on aviation, and later constructed and served as test-pilot of a 600-horsepower airplane for the Austrian army. However, after the German defeat in World War I and his being classified as Jewish by the Nazi government, he fled Germany and eventually landed a staff position at Harvard University in 1939. This same year he introduced the now widely discussed "birthday problem."[8] The "birthday problem" has since gained greater popularity and has been discussed by notable mathematicians, columnists, and talk show hosts, including W.W. Rouse Ball[9], Martin Gardner in his 1957 "Mathematical Games" column[10], and Johnny Carson on a 1980 episode of "The Tonight Show". [11]

The famous "birthday problem," or birthday paradox as it is often referred to, asks the following question "how many people are needed in a room so that the probability that there are at least two people whose birthdays are the same day is roughly half?"[12] Before taking a guess, let us consider a few necessary assumptions. First, let us assume that each

---

[8] The Editors of Encyclopædia Britannica, "Richard von Mises," Encyclopædia Britannica, September 27, 2002, accessed 2017.

[9] W. W. Rouse. Ball, *Mathematical Recreations and Essays* (London: Macmillan, 1959).

[10] Martin Gardner, "Paradoxes Dealing with Birthdays, Playing Cards, Coins, Crows, and Red-haired Typists," *Scientific American*, April 1957.

[11], "The Tonight Show with Johnny Carson." CornellCast. Accessed 2017.

[12] Edward B. Burger and Michael P. Starbird, *The Heart of Mathematics: An Invitation to Effective Thinking* (Hoboken, NJ: Wiley, 2013).

person has an equal chance of being born on any given day. In other words, the probability that someone is born on any day, such as October 8th, is 1/365 since there are 365 days in a year and a person has an equally likely chance of being born on any day. This leads to another assumption; suppose that we do not have leap years, so there are always 365 days in a year, meaning there is always a 1/365 chance of being born on any given day.

After assuming that there is an equally likely chance of being born on any given day and that there are no leap years, consider what you would anticipate the answer to be. You might think that a room with 367 people guarantees that there are at least two people in the room with the same birthday. This is because there are only 365 days a person can be born on, so there cannot be 367 different birthdays, and therefore at least two must overlap. Then, with 367 people in a room, you think in order to make sure that there is about a 50% chance that at least two of them have the same birthday; you need to put about half of 367 people in a room. So, you would divide 367 by two and decide you would need about 183 people in the room for there to be about a 50-50 chance that at least two people have the same birthday. While this seems to be sound logic, in reality, only 23 people are needed in a room for there to be just over a 50% chance that two people share the same birthday.[13] This is the source of the paradox, and its unexpected solution will be explained with the help of a mathematical concept in probability.

The Monty Hall Paradox

Lets Make a Deal may conjure images of screaming fans in gaudy costumes, but the game show and its original host, Monty Hall, led to a well-known paradox that has continued to perplex even those with the highest of IQs.

[13] Ibid.

The Monty Hall paradox, loosely based on the game show, has a complicated history. Officially, it was first submitted to *The American Statistician* in 1975 by University of California, Berkeley professor Steven Selvin. Later, the problem was in an article in the *Journal of Economics Perspective* by Barry Nalebuff and in a 1989 issue of *Bridge Today* by Phillip Martin. However, it was not until a controversial *Parade Magazine* article in 1990 that the paradox rose to fame. Marilyn vos Savant, known for being listed in the *Guinness Book of World Records* for having the "World's Highest IQ" and for establishing the "Ask Marilyn" column in *Parade,*[14] responded to a reader's questions regarding the obscure Monty Hall Problem. In her column, she provided a solution to the paradox that was so counterintuitive that it resulted in her receiving over 10,000 letters, many written by scholars and Ph.Ds., proclaiming "she blew it" and even one saying there was "such a thing as female logic." Through challenging people to complete their own experiments, she eventually convinced 56% of the general public and 71% from academic institutions of her solution.[15] Though much of the dispute has subsided, the paradox and its baffling solution, after the havoc they wreaked on the mathematics and statistics community, remain infamous.

The paradox is as follows. Suppose you are selected as a player on a game show and the host places you in front of three doors. Behind two of the doors are donkeys and behind the third door is a car. The host then asks which door you would like to choose and you select one of the three doors. After you select a door, the host, who knows what is behind all three doors, opens one of the two doors you did not choose and reveals a donkey. You

[14] Zachary Crockett,"The Time Everyone "Corrected" the World's Smartest Woman," Priceonomics, August 2016, accessed 2017.
[15] Marilyn vos Savant,"Game Show Problem," Marilynvossavant.com, accessed 2017.

are then asked if you would like to keep the door you selected or switch to the other remaining door. Assuming you want to win the prize, should you keep the door you first selected, switch, or does it matter?

At first glance, the question appears to be rather simple. If you are given a choice of three doors where you do not know what is behind them, then it would be logical to assume you have a one in three chance of picking the door with the prize behind it. When the host reveals a donkey behind one of the doors you did not choose, it then seems as though your options have been narrowed to two and you should have a 1 in 2 chance of winning with either door, making it irrelevant whether you choose to switch. However, this is where the problem becomes a paradox because once the host reveals a donkey behind a door you did not choose, the odds your door and the remaining door have the prize behind them are no longer the same. [16] The key to understanding this lies in the concept of conditional probability.

No matter how the problem is considered, who is providing an explanation, or the way that the conditional probability and Bayes' rule are manipulated, the solution remains just as counterintuitive, but always the same. The probability of winning is higher if you choose to switch.

While the Monty Hall Problem was first presented in 1975, two other puzzles that were particularly similar predated it. The first of these was Joseph Bertrand's 1889 box paradox. In Bertrand's scenario, three boxes are presented: one with two gold coins, one with two silver coins, and the last with one silver and one gold coin. Suppose a participant draws one gold coin from a box. The question then is what is the probability that the other

---

[16] Eric W. Weisstein, "Monty Hall Problem," WolframMathWorld, accessed 2017.

coin in that box is gold. The answer to this is calculated much the same way as in the Monty Hall paradox and yields the same result. However, unlike vos Savant, Bertrand was celebrated for his findings. After Bertrand, came Martin Gardner's 1959 paradox entitled the Three Prisoner's Problem. Statistically, the scenario is identical to the Monty Hall paradox and Bertrand's box paradox. In his solution, Gardner wisely and perhaps presciently acknowledged "in no other branch of mathematics is it so easy for experts to blunder as in probability theory." [17]

Prosecutor's Fallacy

The Prosecutor's fallacy has plagued the court system likely since the beginning of law itself. However, it was William Thompson and Edward Schumann who put a name to it in their 1987 paper entitled "Interpretation of Statistical Evidence in Criminal Trials: The Prosecutor's Fallacy and the Defense Attorney's Fallacy."[18] The paper mentions a discussion one of the authors had with a Deputy District Attorney about incidence rate statistics. The experienced prosecutor was of the belief that you can determine the probability of a defendant's guilt by subtracting the incidence rate of a "matching" characteristic from one.

In other words, consider a trial that you have witnessed, either in person or from a TV show such as *Law and Order*. Suppose the prosecutor who is trying to prove that the defendant is guilty receives DNA evidence, likely at the last possible second if you are indulging in a crime show. A lab evaluated the blood at the crime scene and found that

---

[17] John Tierney, "Behind Monty Hall's Doors: Puzzle, Debate and Answer?," The New York Times, July 20, 1991, , accessed 2017.

[18] William C. Thompson and Edward L. Schumann, "Interpretation of Statistical Evidence in Criminal Trials: The Prosecutor's Fallacy and the Defense Attorney's Fallacy," *Law and Human Behavior* 11, no. 3 (1987): doi:10.1007/bf01044641.

there was a match between it and the defendant's blood sample. However, the lab also reports that there is a 1/1000 probability of a random match; this is referred to as the incidence rate. The prosecutor reasons from this report that there is a 1/1000th, or 0.1% chance that the DNA came from someone else other than the defendant, so there must be a 99.9% chance that the defendant is the perpetrator and is guilty since one minus 0.1% is 99.9%.

This reasoning however is incorrect because it bases the probability of guilt entirely on one piece of evidence and does not account for other evidence in the case. In addition, it makes the mistake of considering the probability that a defendant is guilty given the evidence, instead of the probability of some evidence given the defendant is guilty. This is a mistake because the probability that the defendant is guilty given the evidence requires that you know the probability that a defendant is guilty. The accused is either guilty or not and there is not a random probability that may be assigned to this. This would only be relevant if the defendant was selected at random from some population that included the guilty party.[19] For example, suppose you are the defendant and you robbed a bank. You could not roll the dice and come up with a random probability that you robbed a bank. You either robbed the bank or did not, and it was not by random chance. The only time this sort of scenario would apply is if there was a group with the robber in it and an officer decided to randomly arrest a member of the group and call them the robber. Then, the odds that this person was the actual robber would be meaningful.

The prosecutor's fallacy represents an error in logic that occurs when conditional probabilities are confused. Conditional probability and the theorem behind it will be

[19] Philip B. Stark, "Probability: Axioms and Fundaments," Statistics at UC Berkeley, accessed 2017.

explained in the next section. This will then lead to a better understanding of how this is an error and, later, why this error could result in wrongful convictions in the courtroom.

Base Rate Fallacy

Suppose you live in a city with two different cab companies. One of the companies drives blue cabs and the other drives green. The blue company is more successful and has 85% of the cabs in the city, while the green cab company operates the remaining 15% of the cabs. Unfortunately, one night a cab was involved in a hit-and-run accident. You witness the tragic event and tell police that you are sure the cab was green. The case is brought to court and you are tested on your ability to distinguish between blue and green cabs during nighttime conditions. After being tested, it is discovered that you were able to identify the color of the cab correctly about 80% of the time, but were guilty of confusing it with the other color 20% of the time. What are the odds that the cab in the accident was actually green as you claim?[20]

While you find yourself upset that you could not determine the color of the cab correctly every time when tested, you sit down to consider the facts and come to the conclusion that there is an 80% chance that the cab in the accident was green. However, if you came to this conclusion then you exhibited the base rate fallacy, which is the fallacy of allowing indicators to dominate base rates in your probability assessments. The base-rates typically come in the form of background data, and in this case, are 85% and 15%, or the given color distribution of the cabs in the city. They can be used to show the true prevalence of an object, event, disease, etc. in a certain population. The 80% and 20% rates

[20] Daniel Kahneman and A. Tversky, "On prediction and judgment," *Oregon Research Institute Bulletin*,1972.

of your color detection, on the other hand, represent indicant or diagnostic information. This type of information tends to relate to the specific object or event being discussed, such as the hit-and-run cab. It has been discovered in a number of experiments that people are inclined to dismiss the base-rate information and place a higher value on the diagnostic information, as you may have done in this question when you assumed that the odds are 80% when they are actually 41%.[21] Among those who have warned against and popularized this fallacy, are Meehl and Rosen. Meehl and Rosen, in their 1955 paper, wrote about the dangers of psychologists evaluating patient's test results on diagnostic information alone, rather than taking into account base-rates, and other relevant information, such as costs and goals. They found that clinicians were no less confident or skeptical when a patient's test results yielded a rare result, such as 'suicidal.'[22] Although the base-rates can be easily dismissed, it is clear in this situation how they can have an important impact.

Simpson's Paradox

Simpson's paradox occurs when there are two different categorical variables that when considered together are qualitatively different than when considered apart. The history of the Simpson's paradox is also somewhat of a paradox, as the statistician Edward Simpson did not discover it and described a slightly different phenomenon in his 1951 paper on association paradoxes, and yet the paradox bears his name. Association paradoxes, of which Simpson's paradox is one, and the discovery of them, can be attributed

---

[21] Maya Bar-Hillel, "The Base-Rate Fallacy in Probability Judgments," *Acta Psychologica* 44, no. 3 (1980): , doi:10.1016/0001-6918(80)90046-3.

[22] Paul E. Meehl and Albert Rosen, "Antecedent probability and the efficiency of psychometric signs, patterns, or cutting scores.," *Psychological Bulletin* 52, no. 3 (1955): doi:10.1037/h0048070.

to the British statistician Karl Peason. In 1899, Pearson showed that marginal and partial associations between continuous variables might produce different results and lead to incorrect correlations. In his study, Pearson demonstrated that male skulls from the Paris catacombs had a 0.09 correlation in their lengths and breadths. However, the length and breadth of the female skulls produced a negative 0.04 correlation. Then, when the samples were combined, the joint correlation was found to be 0.2. From this Pearson determined that skull length and breadth were uncorrelated for males and females separately and positively correlated for males and females jointly. After Pearson identified that the paradox could occur with continuous variables, George Udny Yule, a British statistician, reported in 1903 that the same paradox could occur with categorical variables.[23]

The paradox is best understood through an example. One such example occurred in a gender discrimination suit against the University of California, Berkeley. Berkeley had 8,442 male applicants and 4,351 female applicants for graduate school in the fall of 1973. Of the male applicants, 44% were admitted, while only 35% of the female applicants were admitted. This would suggest that gender discrimination occurred. However, when researchers looked more closely at the admission statistics, they found that this was not the case. In fact, it was quite the opposite in certain departments. This suggests a paradox because how could there be a higher percentage of males admitted than females if females had higher admission rates than males in numerous departments?

[23] Bruce W. Carlson, "Simpson's paradox," Encyclopædia Britannica, August 31, 2016, accessed 2017.

Expected Value Theorem Paradoxes

Allais Paradox

Suppose you find yourself in a gambling mood and decide to enter a casino with a few friends. In this particular casino, you are given the choice between two different gambles. The first is gamble A, which 100% guarantees you $100 because this is a very generous casino. The second is gamble B, which is a bit more complicated. Gamble B gives you a 10% chance of winning $500, an 89% chance of winning $100, and a 1% chance of receiving nothing.[24] Which do you choose? Gamble A has the benefit of no risk, but gamble B gives you a small chance of receiving a higher payout. Your immediate thought is to go with gamble A because you believe you cannot beat a guaranteed win.

After successfully deciding on your first gamble, the casino decides to offer you yet another two gambles to decide between, gamble C and gamble D. Gamble C gives you an 11% chance of receiving $100 and an 89% chance of receiving nothing. On the other hand, gamble D gives you a 10% chance of receiving $500 and a 90% chance of receiving nothing.[25] You become a bit concerned with all of the percentages, but notice that gamble D gives you only a 1% lower chance of winning but has a much higher payout. You carefully weigh your choices and then decide to take the risk and go with gamble D since it has the possibility of a much higher payout.

Before officially deciding to make gamble A and gamble D, you decide to check with your friends who already have their winnings to learn which of them won the most and what gambles they chose. You discover that the majority of them, whether they won or lost,

---

[25] "Allais Paradox," Policonomics, 2012, accessed 2017.
(Gamble A,B,C,D and the numbers behind them all come from this source)

made the same gambles that you plan to make, but that the winner chose gamble B and gamble D. You think maybe this was just luck, but should you switch to gamble B?

According to expected value theory, the short answer is yes; you should choose gamble B and gamble D. The theory claims that people make decisions by picking the option that will provide the highest expected value to them, where value refers to a monetary amount.[26] In this case, the numbers reveal quite obviously that gamble B and gamble D have the highest expected values. You considered yourself logical and know that the majority of your friends made the same decisions you did, so it seems unreasonable that this theory claims gamble B has a higher value than A. How can winning 100% of the time not be the best gamble to make? In other words, why would the average and reasonable person who can assign value to decisions make a decision with a lower value?

This contradicting notion that people might go against expected value theory, despite expected value theory rationally claiming people make decisions based on what will give them the highest expected value represents a paradox known as the Allais paradox.

The Allais paradox was first mentioned in a 1953 French paper entitled "Le Comportement de l'homme rationnel devant le risque: critique des postulats et axiomes de l'école américaine." The paper was written by French economist and physicist, Maurice Félix Charles Allais. In his paper, Allais successfully proved that the Allais paradox failed decision theory due to its violation of expected value theory. Expected value theory may succeed in situations where a person is presented with risk one way and fail in situations where a person is presented with risk in another. The theory is successful when people are

[26] Rachael Briggs, "Normative Theories of Rational Choice: Expected Utility," Stanford Encyclopedia of Philosophy, August 08, 2014, accessed 2017,

in an uncertain environment where risk is abundant. On the other hand, when there is an opportunity for a certain positive outcome, then the theory fails because the average person tends to favor situations with no risk.

The idea that Allais developed about people performing differently in uncertain environments also served as a key factor in helping Allais prove that effectively allocating risk may lead to a more optimal allocation of resources. This then contributed to him later winning the Nobel Prize in Economic Sciences for his work with the theory of markets and his studies on the most effective methods for consuming resources. In addition to winning the Nobel Prize, Allais and his work inspired future economists and psychologists, including Daniel Kahneman, to delve further into the field of behavioral economics.[27]

St. Petersburg Paradox

Another paradox that explores the idea of expected value theory is that of the St. Petersburg Paradox. The St. Petersburg paradox was first discovered by Nicolaus Bernoulli, a Swiss eighteenth-century mathematician. However, it was published by his brother, Daniel Bernoulli, in the *St. Petersburg Academy Proceedings* in 1738,[28] which is what gave it its name. The paradox can best be understood through the flipping of a coin.

Imagine that you decide to go back to the casino and are given the opportunity to play the St. Petersburg game. You are given a fair coin and told to flip it until you get a tails. If the number of flips that it takes to get a tail is equal to n, then n flips will you give $\$2^n$ worth of prizes. For instance, if you get a tail on the first flip, then you would win $\$2^1=\$2$. If you were to get a tail on the second flip, then you would win $\$2^2=\$4$ and so on. If you were

[27] "Maurice Allais," Policonomics, 2012, accessed 2017

[28] Robert Martin, "The St. Petersburg Paradox," Stanford Encyclopedia of Philosophy, November 04, 1998, accessed 2017.

to keep flipping heads, then the game could potentially continue on into infinity. This poses a problem because if the game could be played an infinite number of times, then the prizes could also grow to be arbitrarily large.

If the expected value theory is applied, then each flip's expected payoff turns out to be $1. If the game is played infinitely, then there is a chance a person could win an arbitrarily large sum of money. Even if there were a high finite entry fee for playing the game, it would still be worth the gambler paying it because they have the chance to win an extremely large amount of money. However, it does not seem reasonable that someone would be willing to pay an extraordinarily high entry fee for such a game. There must be a cutoff point where a person no longer wishes to enter into such a game because the price to play it is too high. This is the paradox because while it seems logical that the game could be played an infinite number of times and give a player a finite but exceedingly large number of dollars despite any high but finite entry fee, it seems irrational that a person would play the game no matter how high the price.[29]

Newcomb's Paradox

Suppose you are walking along and suddenly run into a Wise being, which we will refer to as W. W tells you that it has placed $1000 in a box labeled box A. Next, W explains that it has another box, box B, with contents of either $1 million or nothing. W then tells you that you can take the contents of box B only, or take the contents of both A and B. You think this sounds too good to be true so you decide to figure out the catch. W explains that if an algorithm predicts that you will only take box B, then box B will contain $1 million.

[29] Ibid.

However, if the algorithm predicts that you will choose box A and box B, then box B will contain nothing.

Since W uses this algorithm and he is a wise being, you can assume that the algorithm never fails. It will make its prediction and then you will make your choice. However, at the time of your choice, you do not know the algorithm's prediction, so which box or boxes do you choose?

You decide to consult game theory before making your choice. However, this leads you astray, because you come up with two rational answers that contradict each other. The first approach states that W has designed a prediction algorithm whose answer will always match what you choose to do. By this logic, if you choose both boxes, then you will always receive $1,000 since W knew to put nothing in box B. Alternatively, if you choose only box B, then you will receive $1 million because the algorithm will have accurately predicted that you would only choose box B. In this approach, you should then always choose only box B.[30]

The second approach states that you should take both boxes. Your choice occurs after W has made its prediction and you believe that you have free will, so you can choose whatever choice you want and it will be independent from the prediction W made. With this logic, if W thought you would take boxes A and B, then taking both is guaranteed to give you $1,000. This is because while W would have put nothing in box B, you will still win the $1,000 in box A no matter what. However, if W predicted that you would only take B, then taking both boxes would give you $1,001,000. This is still better than only choosing

[30] The previous description of the paradox and the approaches used to solve it were taken from the following source.
Gregory Benford and David Wolpert, "What Does Newcomb's Paradox Teach Us?," *SSRN Electronic Journal*, 2010, doi:10.2139/ssrn.1381295.

box B because in this case, if you only chose box B, then you would receive $1 million. Therefore, by this method, you should always take both boxes.

Since both of these answers seem logical but contradicting, you can see how this is a paradox. It is referred to as Newcomb's paradox and was created by William Newcomb in 1960.[31] Interestingly enough, Newcomb never published the paradox, but did discuss it at length with philosophers and physicists, including Robert Nozick and Martin Kruskal, and later with Scientific American columnist Martin Gardner.

Nozick featured the paradox in his 1969 paper, in which he suggested, "To almost everyone, it is perfectly clear and obvious what should be done. The difficulty is that these people seem to divide almost evenly on the problem, with large numbers thinking the opposite half is just being silly."[32] In addition, he discussed how the two accepted principles of game theory appear to conflict. The expected value theorem takes into account the probability of each outcome and states that you should take box B only. On the other hand, the dominance principle states that in situations where one strategy is always better than the other strategies no matter what the other players do, then you should pick that strategy. In Newcomb's paradox, regardless of what box B contains, you will always receive $1,000 more if you take both boxes than if you just take box B. Because of this, the dominance principle claims that you should always take both boxes. In addition, it is important to note that Nozick's approach presumed that W's predictions were of high accuracy, but were not certain. He also excluded the idea of backward causation, or the idea that for predictions made in the present to be perfectly determined by events in the future

---

31 Martin Gardner, "Reflections on Newcomb's Problem: a Prediction and Free-Will Dilemma," *Scientific American*, March 1974.

32 Robert Nozick, "Newcomb's Problem and Two Principles of Choice," *Essays in Honor of Carl G. Hempel*, 1969, doi:10.1007/978-94-017-1466-2_7.

means that the future causes past events.[33] This eliminates the concept of time travel, which was discussed by Newcomb in a paper around the same time he created the paradox.[34]

Along with Nozick, Martin Gardner also wrote about the paradox. In his long running "Mathematical Games" column for *Scientific American*, Gardner presented the paradox twice, claiming that the second time generated more mail than any of his other articles.[35] Gardner, as Nozick did, discussed the paradox at length with Newcomb and offered both possible solutions, providing arguments for each course of action. Further, he concluded with the question "Can it be that Newcomb's paradox validates free will by invalidating the possibility, in principle, of a predictor capable of guessing a person's choice between two equally rational actions with better than 50 percent accuracy?"[36]

After contemplating these two lines of reasoning, let us consider what solution Newcomb believed to be correct. William Benford, an author of "What Does Newcomb's Paradox Teach Us?" worked with Newcomb and also often discussed the paradox with him. Benford claims that Newcomb created the paradox to test ideas, but believed in the second solution, or the idea that you should just take box B because why fight a God-like being?[37]

Regardless of which you would choose, it is important to understand the mathematical approaches to both solutions, as the math may change your mind once again.

---

[33] Ibid.
[34] Ibid. 30.
[35] Ibid.
[36] Martin Gardner, "Free Will Revisited, With a Mind-Bending Prediction Paradox by William Newcomb,"*Scientific American*, July 1973.
[37] Ibid. 30.

Regression To The Mean

Much of statistics came to be in the late 19th and early 20th centuries as a result of the study of heredity and scientists' fascination with the likelihood of certain features in one generation being passed down to the next. In the 1880's, Francis Galton led a heredity study on the adult heights of parents and their children in an effort to discover the extent to which height is inherited. The data collected was similar to his protégé, Karl Pearson, which showed that the twenty tallest fathers were on average 6.2 inches taller than their generation's average height, while the sons were on average only 2.8 inches above their generation's average height. In addition, in a study of the twenty shortest fathers, the fathers were on average 6.9 inches below their generation's average height, while the sons were only about 3.3 inches shorter than their generation's average height. In other words, the sons of the tallest fathers were taller than average, but not by as much as their fathers were and the sons of the shortest fathers were shorter than average, but not by as much as their fathers were. Galton called this effect of height getting closer to the average a "regression toward mediocrity," where mediocre was thought of as average.[38] Let us discuss a different example of this phenomenon.

Consider the last sport or game that you played. Were you in a slump or were you lucky enough to be on a winning streak? If you found yourself in either of the extremes, then like it or not, it probably did not last or will not continue the next time you play. This comes as a result of the regression to the mean phenomenon.

The regression to the mean phenomenon occurs when an extreme outcome is followed by an average outcome and where the game or event taking place is governed, at

---

[38] Daniel Kaplan, *Statistical Modeling: A Fresh Approach* (Lexington: S.n., 2013).

least in part, by chance. The problem occurs when it is assumed that after an extreme outcome has occurred, a change has taken place and this extreme will continue.[39] In other words, just because you were losing does not mean you will continue to lose, and, in fact, you will likely go back to playing how you usually do. Similarly, if you are on a winning streak, then this does not mean that you will continue winning, but likely that you will eventually go back to playing how you were previously. While it may be reassuring to find out that losing a few rounds of a game does not mean that your talent has diminished, it is less exciting to accept that your winning streak may not be due to your increased ability.

The same holds true for many of the famous athletes that grace the pages of our magazines. The *Sports Illustrated* curse has become infamous as those who have dared to pose for the popular magazine have found that their previous game-winning plays and/or seasons have quickly transformed into losses and injuries after the cover featuring them hit the shelves. [Insert Heart of Math info] The curse has supposedly claimed many victims, including most notably, the Boston Red Sox and Chicago Cubs, Tom Brady, and Michael Spinks.

In October of 2003, the Boston Red Sox and Chicago Cubs were featured on the cover of *Sports Illustrated* as they were touted to win their respective league's championship series. However, as luck would have it, New York Yankee, Aaron Boone, made a walk-off home run to defeat the Sox, and the Cubs fell to the Florida Marlins. In the case of New England Patriots' quarterback, Tom Brady, *Sports Illustrated* chose the player to be on the cover for the 2008 NFL season preview. In the photo, Brady is seen stretching with his knee exposed, which proved to be ironic when during week one, he tore his ACL

[39] David M. Lane, "Regression Toward the Mean," OnlineStatBook, 2016, accessed 2017.

and MCL and missed the entire season. Another casualty of the curse was boxer Michael Spinks. Spinks found himself on the cover in June of 1988 just before his fight against Mike Tyson. Despite the cover containing the words "Don't Count Me Out," it only took 91 seconds for Spinks to be knocked out by Tyson.[40] Countless others seem to give proof to this legend of the *Sports Illustrated* Curse, but in reality, the athletes and their fans have failed to consider the regression to the mean phenomenon that is unavoidable.

The curse not only suggests that athletes after playing extremely well and experiencing big wins will likely return to playing at their average level, but also that those experiencing big wins will most assuredly later be afflicted with extreme losses. While it is true that athletes will most of the time play at whatever level is average for them with some highs and some lows, it is not true to assume that highs must be followed by lows or vice versa, so that a player stays playing close to average.[41] Consider you are playing a game of darts and hit a bullseye. Have you hit a bullseye the exact same number of times that you have thrown a dart and not hit the board at all? It is unlikely and you would not expect that for every time you have missed the board you will hit a bullseye. The same holds true for athletes at the top of their game. Curse or no curse, just because a team has made it to the league's championship series, a player experienced a winning streak the previous season, or a boxer is undefeated in their professional career, does not mean that they will experience their greatest win or suffer an incredible loss or injury or series of losses or injuries. Instead a star athlete, or an amateur dart player, will likely play close to how they usually do.

---

[40] Nicholas Parco, "A Look at Victims of the Sports Illustrated cover jinx," NY Daily News, March 23, 2016, accessed 2017.

[41] A. G. Barnett, "Regression to the mean: what it is and how to deal with it," *International Journal of Epidemiology* 34, no. 1 (2004): doi:10.1093/ije/dyh299.

The sophomore slump is another misunderstood occurrence similar to the *Sports Illustrated* curse. Players with the highest batting averages in their first year more often than not experience lower averages in their second year. Many attribute this to a slump. While it is true that their averages are often lower, this is not primarily due to their decreased ability, but because of the regression to the mean. They are more likely to play closer to average in their next season, since baseball is based not only on ability but luck. While ability may be a controlled element that is relatively static, there is an element of chance or luck that may vary from game to game and season to season.

The regression to the mean may be explained mathematically through the concept of the bivariate normal distribution. This probability distribution takes in to account two random variables and their means, variances, and covariance. The math behind this distribution and its relation to the regression to the mean phenomenon will be explained further.

SECTION IV: THE MATHEMATICAL CONCEPTS BEHIND THE SOLUTIONS

Basic and Conditional Probability

Basic and Conditional Probability: Axioms, Rules, and Theorems

Behind solutions to many of the paradoxes and fallacies are mathematical concepts in probability theory. A brief description of the basics may aide in the understanding of both the paradoxes and fallacies that have previously been discussed and those one might run into in a real world setting.

Before covering the more challenging concepts in probability, it is first vital to comprehend a few of the rules probability, as a whole, must satisfy. In mathematics, these are referred to as axioms. The first of these is that the probability of every event occurring must be greater than or equal to zero. In other words, there cannot be a negative chance of something happening. The second axiom states that the probability of the entire outcome space is 100%, where the outcome space contains every possible outcome. Another way of saying this would be if you flip a coin, then there are two possible outcomes that could occur: you could flip a heads or you could flip a tails. Each outcome has a 50% chance of happening and there are two possible outcomes, so there must be a 100% chance that the coin will land on heads or tails. The third axiom states that if two events are disjoint, or mutually exclusive, meaning they cannot both happen at once, then the probability that either of the events occurs is the sum of the probabilities that each occurs[42]. In other words, if you were trying to find the probability of rolling a two or a five with a fair die, then you would add the chance of rolling a six-sided die and getting a two (1/6) to the probability of rolling a five (1/6) to get a 2/6 or 1/3 chance of rolling either a two or a five.

[42] Technical Definitions of Axioms from the following source: Ibid. 19.

All of probability builds on these three axioms, and one consequence of them that will serve as a particularly useful tool in the solving of several of the paradoxes is the complement rule. The complement rule states that the probability that an event occurs is always equal to 100% minus the probability that the event does not occur.[43] For example, if there is a bag of three marbles where one is red, one is blue, and one is green, then there is a 1/3 chance of drawing a red marble. The probability of not drawing a red marble is the sum of the probabilities of the other outcomes, so the probability of drawing a blue (1/3) plus the probability of drawing a green (1/3), which is 2/3. If we did not previously know the probability of drawing the red marble, then we could find it by using the complement rule. The rule states that the probability of drawing the red marble is equal to (3/3, or 100%, or 1) minus the probability of not drawing a red marble, or (2/3). Then, 1-(2/3)=(1/3), which is the probability of drawing a red marble. In certain situations, it may be easier to first find the probability that an event does not occur, and then subtract this from 100% to find the probability that it does occur.

Another important concept in probability is that of independence. Independence and mutual exclusivity are often confused, and it is important to know the difference. As discussed above, mutually exclusive events are events that cannot both occur in the same trial, so if you know that event B happened, then it implies event A did not happen. In other words, if you flipped a coin and it landed on heads, then you know that the coin did not also land on tails. In addition, with mutually exclusive events, the probability of event A or event B happening can be found by adding their individual probabilities. Independent events on the other hand, can happen in the same trial, except if at least one of them has a probability

[43] Ibid.

of zero. Independent events occur when learning that one event happened does not give you any information about whether the other happened.[44] For instance, if a round of a game requires you to flip a coin and roll a die, then flipping a tails does not reveal what number you will get when you roll the die. Further, the probability of event A **and** event B occurring if events A and B are independent can be found by multiplying the individual probabilities. This means that the probability of flipping a tails (1/2) AND rolling a three (1/6) would be (1/2)*(1/6), which equals 1/12. The probability of A or B where A and B are independent events is known to be less than the sum of the individual probabilities, unless at least one of the events has zero probability. This means that the probability of flipping a tails OR of rolling a three would be less than (1/2) + (1/6), or 2/3.

Independence plays an important role in conditional probability. Conditional probability measures the probability of event A occurring given that event B occurred. If the two events are independent, then the probability that event A occurred given that event B did will be whatever the probability of event A occurring is, since the two events are not affected by one another. In other words, the probability of flipping a tails, given that you rolled a three, is not impacted by the number that you rolled, so it will still have the same individual probability of occurring, which is ½ since you have a one in two chance of flipping a tails. However, if the two events are not independent, then the conditional probability becomes more complicated to calculate. For example, if you have a deck of cards, then the probability that you draw a card in a certain suit is 13/52. Suppose that you draw a heart. Then, without putting this card back in the deck, if you draw again, then the probability of getting another heart will be 12/51 since there is one less heart and one less

[44] Ibid.

card in the deck. This represents conditional probability since the probability of drawing a second heart is impacted by whether or not you drew a heart the first time.

In cases where the events are not independent and the results are a bit less intuitive than the previous example above, then Bayes' Rule is used to find the conditional probability. Bayes' rule states that the probability that event B occurs given that A occurred is equal to the probability that event A and B occurred divided by the individual probability that event A occurred.[45] For example, suppose 80% of your friends live in Texas and 60% live in Texas and are Longhorn fans. What is the probability that your friends are Longhorn fans given that they live in Texas? If you apply Bayes' rule, then you would take the number of friends that live in Texas AND are Longhorn fans (60%) and divide this by the number of friends that live in Texas (80%). The probability that your friends are Longhorn fans given that they live in Texas is then 60/80, or 75%.

By the multiplication rule, Bayes' theorem may also be rearranged to calculate the probability that event A and event B occurred. The probability that event A and event B occurred is equal to the probability that event A occurred multiplied by the probability that event B occurred given that A occurred.[46] For example, suppose that there are seven marbles in a bag. Three of them are white and four of them are orange. What is the probability that you first drew a white marble and draw an orange marble second? If you apply the theorem, then the probability that you choose a white marble first and an orange marble second can be calculated by first finding the probability that you choose a white marble first. This is equal to 3/7 since there are three white marbles and seven total marbles. This then is multiplied by the probability that you choose an orange marble

---

[45] Ibid.
[46] Ibid.

second, given that you chose a white marble first. This is equal to 4/6 since there is still the same number of orange marbles, but there is one less white marble and therefore one less marble in the bag. Thus, the probability that you choose a white marble first and an orange marble second is equal to (3/7)*(4/6), which is 12/42, or 2/7. Bayes' Rule and its different forms are particularly helpful in understanding the solutions and explanations behind many of the paradoxes and fallacies that have been discussed.

The Birthday Paradox Solution

The Birthday paradox, which was the first paradox discussed, asks how many people must be in a room so that the probability that there are at least two people whose birthdays are the same day is roughly one-half. Many of the probability concepts that were just explained will be useful in forming the solution to this paradox.

Given the assumptions made previously, it is first necessary to find the probability that two people share the same birthday. In other words, what is the probability that the second person chosen has the same birthday as the first person? If we were to tackle the question this way, then we would first consider the number of possible pairs of dates there are for two people. If there are 365 possibilities for the first person and 365 possibilities for the second person, then there are 365*365=133,225 possible pairs of birthdays. However, there are only 365 times where the first person has the same birthday as the second person, so the odds that two people have the same birthday are 365/(365*365), which equals 1/365.

Instead of continuing this and finding the probability that two out of three people have the same birthday, then two out of four people have the same birthday, etc., it may be easier to consider finding the complement, or the probability that two people do not have the same birthday, and then applying the complement rule to find the probability that they do. Finding the probability that two people do not share the same birthday requires that we consider there are, as before, 365*365 different pairs of birthdays. However, this time it is key to think about how many of these possible outcomes have a pair with two different birthdays. Since there are 365 possible days that the first person's birthday might fall on, then there are 364 days that the second person's birthday could fall on and still have a different birthday from the first person. Therefore, there are 365*364=132,860 pairs that have two different birthdays. The probability of two people having different birthdays is then (365*364)/(365*365), which is equal to 364/365 or 0.9972… When the complement rule is applied to find the probability that two people have the same birthday, then 1-(364/365)=(1/365) which is equal to 0.00273. Thus, once again it is confirmed by a different method that there is a 1/365 chance that two people have the same birthday.

Let us apply this second method to a room with a few more people in it. Suppose there are three people in a room. Then, as we calculated before, we must first find the number of possible triples of dates there are for three people. If there are 365 possible dates for the first person, 365 possible dates for the second person, and 365 possible dates for the third person, then there are 365*365*365= 48,627,125 possible triples of dates. After this we have to find the number of dates such that all three people have different birthdays, and then we can apply the complement rule. The first person could have a birthday on any day, so they have 365 different possibilities. The second person could have a birthday on any day of the year other than the date of the first person's birthday, so they have 364 possibilities. Then, the third person can have a birthday on 363 days since their birthday can fall on any day other than the first person's birthday and second person's birthday. Thus, there are 365*364*363=48,228,180 possible triple dates where all three people have different birthdays. Therefore, there is a (365*364*363)/(365*365*365) probability that all three people have different birthdays. We can then apply the complement rule to find the probability that the opposite is true. This is equal to 1-(365*364*363)/(365*365*365)= 1-0.9917= 0.0082, which is the probability that at least two of the three people have the same birthday.

0.0082 is not equal 0.5, which is the probability that we are striving for, so we must continue to add people to the room until we reach this desired probability. While 0.0082 is not close to 0.5, it is quite a leap from, and in fact almost three times greater than, 0.00273, which was the probability that two people in a room have the same birthday. With this in mind, we can see that the probability increases dramatically with only one person added to the room. Because of this, we may get to the answer faster than we think.

In order to speed up the process, suppose we increase the number of people in the room by 5 each time. Then, the following table[47] of values would be created.

Number of People in The Room	**Probability of at least two sharing the same birthday**
5	0.027…
10	0.116…
15	0.252…
20	0.411…
25	0.568…
30	0.706…
40	0.891…
50	0.970…
60	0.994…
70	0.9991…
80	0.99991…
90	0.999993…

From this table, we can surmise that the probability reaches just over 0.5 in a room with between 20 and 25 people in it. If we were to calculate the probabilities of 21,22,23, and 24 people in a room, then we would discover that 23 people are required in a room for the probability to reach 0.5. In addition, the table shows us that in a room of 50 people there is a 97% chance, and that in a room of 90 people, there is almost a 100% chance that two people share the same birthday.

From the math behind this paradox, we can see that there are times where our logic fails us. This may be particularly true in cases where small probabilities play a key role. However, with a careful understanding of probability and the ability to consider the complement, or opposite, of a situation we may be able to easily calculate and understand solutions to problems such as the birthday paradox. If all else fails and you still remain unconvinced, try the experiment for yourself. If you find yourself in rooms with fellow students, coworkers, or friends in varying numbers, then see if they will volunteer their birthdays. If you perform it enough times, you will end up with the same result: with 23 people in a room, there is a better than even chance at least two of them share the same birthday. [48]

The Monty Hall Paradox Solution

The Monty Hall paradox requires that we use Bayes' theorem, as well as the complement rule, which we recently employed in the birthday paradox. First, we must examine the different scenarios that could occur after picking a door and then calculate using Bayes' Rule to discover which door would have a better probability of having the car.

---

48 The explanation and table provided to explain the birthday paradox came from the following source: Ibid.12.

Then, we may use the complement rule to determine the probability of winning if we switch doors, which will then tell us whether or not we should switch doors.

Suppose the player chooses door 1, then there are three possible cases that could occur. The first is that there could be a donkey behind door 1, a donkey behind door 2, and a car behind door three, and so the host will then open door 2. The second case is that door 1 could have a donkey behind it, door two could have the car, and door 3 could contain the donkey, and the host would then open door 3. The final case occurs when door 1 has the prize behind it, and therefore doors 2 and 3 have donkeys behind them, meaning the host may open either of these doors. If you choose not to switch doors, then you want the final case, where you chose the correct door initially. With this in mind, then you want to calculate the conditional probability that the last scenario occurs given that the host opens door 2 or door 3. If the host opens door three, then door one, your door, and door two are left. It is important to note that because the prize must be behind one of these two doors, the probabilities of the two doors having the prize must sum to 100%. Below is a list of the cases.

	#1	#2	#3	host's action
case 1	donkey	donkey	car	opens #2
case 2	donkey	car	donkey	opens #3
case 3	car	donkey	donkey	opens #2 or #3

Using Bayes' Rule, the conditional probability that the third case occurs (event B) given that the host opens door 3 (event A) may be found by dividing the probability that case 3 happens and door 3 is opened (event A and B) by the overall probability that door three is opened by the host (event A). From the list of cases, door 3 is opened in case 2 and in case 3, so to calculate door 3's overall probability of being opened, its probability in each of these cases must be found and then summed. The probability of case 2 and door 3 opening is equal to 1/3 since there is a 1 in 3 chance of case 2 happening. The probability of case 3 and door 3 opening is equal to the probability of case 3 multiplied by the conditional probability of door 3 opening given case 3 occurs. This is a modified version of Bayes' Rule that occurs when the multiplication rule is applied. This is equal to 1/3 since there is a 1 in 3 chance of case three occurring multiplied by 1/2, since in case 3, the odds door 3 and door 2 are chosen by the host are equal because they both have donkeys. This is then equal to 1/6, which can then be inserted into the original conditional probability equation. The probability of case 3 and door 3 opening is then 1/6 divided by the probability of door 3 opening in any of the cases, which is (1/6 + 1/3). This gives the probability that door 1, and therefore any door that you choose arbitrarily at the beginning of the game show, has a 1/3 chance of having the car behind it assuming you do not switch. However, because the probabilities of door 1 and door 2 must sum to 1 since there is a 100% chance the car is behind one of these two doors, then the probability of door 2, or the door you did not choose and the host did not open, will always have a 2/3 chance of winning. This means door 2 has consistently better odds than door 1 in the third case, and that you should **always** switch doors from the one you originally chose. An explanation of these calculations can be seen below.[49] Note that "Pr" refers to the probability of the event in parentheses. The line separating the events is the symbol for "given." The comma is then meant to represent "and."

---

[49] "Conditional Probability, The Monty Hall Problem," Conditional Probability, 2008, accessed 2017.

	#1	#2	#3	host's action
case 1	donkey	donkey	car	opens #2
case 2	donkey	car	donkey	opens #3
case 3	car	donkey	donkey	opens #2 or #3

Now $P(\text{case } 2, \text{open door } \#3) = 1/3$ and

$$P(\text{case } 3, \text{open door } \#3) = P(\text{case } 3)P(\text{open door } \#3|\text{case } 3) = \frac{1}{3} \cdot \frac{1}{2} = \frac{1}{6}$$

Adding the two ways door #3 can be opened gives $P(\text{open door } \#3) = 1/2$ and it follows that

$$P(\text{case } 3|\text{open door } \#3) = \frac{P(\text{case } 3, \text{open door } \#3)}{P(\text{open door } \#3)} = \frac{1/6}{1/2} = \frac{1}{3}$$

An alternate way to think about this is that the only way for the player to lose, given that they always switch is if they choose the correct door initially. There is a 1/3 chance of the player choosing the right door initially, and therefore a 1/3 chance of them losing given that they always switch doors. Thus, there must be a 2/3 chance of winning given that they always switch since the probabilities of each action must add to 1.[50]

Prosecutor's Fallacy: Explanation of Error

Bayes' rule as we have seen can be used to solve several paradoxes and fallacies. Among these is the Prosecutor's Fallacy.

As previously mentioned, Bayes' Rule states that the probability that event B occurs given that A occurred, or Pr (B | A), is equal to the probability that event A and B occurred divided by the individual probability that event A occurred.

$$\Pr(event\ B|\ event\ A) = \frac{\Pr\ (event\ A\ and\ B)}{\Pr\ (event\ B)}$$

[50] Sal Kahn. "The Monty Hall Problem." Khan Academy. Accessed 2017.

The Prosecutor's fallacy occurs when the probability that event B occurs given that A occurred, Pr(B|A), is confused with the probability that event A occurs given that B occurred, Pr (A|B).[51]

As discussed in the description of the prosecutor's fallacy, suppose you are watching a trial where the prosecutor receives DNA evidence from a lab. Assume the lab finds a match between the defendant and the blood sample found at the crime scene. However, the lab also states that there is a 1/1000 chance of a random match occurring. The prosecutor makes the mistake of assuming that there is therefore a 1/1000th, or 0.1% chance that the defendant is innocent and a victim of a random match. By this logic, there must be a 99.9% chance that he or she is guilty. Let us apply Bayes' theorem to show the error.

Let H be our hypothesis about whether the defendant is guilty or not guilty and E be a piece of evidence. Our goal is to find out the probability that our hypothesis is correct, whatever it may be, given the evidence, or Pr(H|E). Applying Bayes' rule, Pr(H|E)=Pr(H and E)/Pr(E).

However, we do not have Pr(H and E), or the probability that our hypothesis is correct and there is evidence, so we must set up another equation to find Pr(H and E) or something that we can substitute for it.

Consider Pr (E| H)= Pr(H and E)/Pr(H). By the multiplication rule, we can see that Pr(H and E) = Pr(E|H)*Pr(H). Thus, we can substitute this in for Pr(H and E) in our original formula.

So, now we have the Pr(H|E)=Pr(E|H)*Pr(H)/Pr(E).

---

51 Ibid.19.

Now, we must consider the overall probability of the evidence. There is the probability of our hypothesis being correct and having evidence, as we just considered. However, there is also the probability of our hypothesis being incorrect and having evidence. Thus, the overall probability of evidence is equal to the probability of a correct hypothesis and evidence plus the probability of an incorrect hypothesis and evidence. As we calculated previously, the probability of the correct hypothesis and evidence is equal to Pr(E|H)*Pr(H). The probability of the incorrect hypothesis and evidence is then equal to Pr(E| not H)*Pr(not H). Thus, Pr(E)= Pr(H and E) + Pr(not H and E)=Pr(E|H)*Pr(H)+Pr(E| not H)*Pr(not H).

If these are combined, then the formula is Pr(H|E)=Pr(H and E)/Pr(E)= Pr(E|H)*Pr(H)/ [Pr(E|H)*Pr(H)+ Pr(E| not H)*Pr(not H)]. Once we have the formula in the correct form, we can begin to consider what numbers we can plug in.

Assuming we believe someone is innocent until proven guilty and therefore that the defendant is as likely as anyone else to have committed the crime on the island before evidence is considered, the probability of our hypothesis is 1/10,000 since there are 10,000 people on the island. Then, we can apply the complement rule to find the probability of not H, which is 1-(1/10,000). The Pr(E|H) is the probability of the evidence given the correct hypothesis, which is the same as the probability of correctly matching a DNA trace, which is equal to 1, since there is no chance of a false negative DNA match. The Pr(E| not H) is equal to the probability of a match in a person who did not have DNA at the crime scene, or the random DNA match probability, 1/1000 as the lab claims.

If we put these probability values into the formula then,

Pr(H|E)= (Pr(E|H)=1)*(Pr(H)=1/10,000)/[(Pr(E|H)=1)*(Pr(H)=1/10,000)+ (Pr(E| not H) =1/1,000)*Pr(not H)=1-(1/10,000)]= (1/10,000)/[(1/10,000) + (1/1,000)*(1-(1/10,000))]=0.090917….., or about 9%.

The probability of getting the correct hypothesis given the evidence is then equal to about 9%. In other words, the probability that we believe the crime scene DNA came from the defendant is then 9% and the probability that we believe the defendant did not have their DNA at the scene is then 100%-9%= 91%. [52]

91%, or the Pr(not H|E), is therefore very different from 1/1,000=0.1%, or Pr(E| not H). Prosecutor's fallacy occurs when a prosecutor observes the DNA statistics and assumes that the probability that the defendant was not the source of the evidence is 0.1%, when in actuality it is 91%. If the prosecutor assumes both are 0.1%, then they may ascertain that the defendant has a 99.9% chance of being guilty, when it is closer to 9% if this is the only evidence presented. It is easy to see how this big of a difference can lead to much higher assignments of guilt and as a result, wrongful convictions.

Base Rate Fallacy: Explanation of Error

The Base Rate Fallacy involves an error in the calculation of Bayes' Rule. It occurs when the probability of an event A or the probability of an event B are ignored in the calculation of Pr(A|B) and Pr(A| not B). This can be seen if we review the cab example. In the example, the blue cab company owned 85% of the cabs in the city and the green cab

---

[52] This scenario and the data behind it was adapted from the following source. Norman Fenton, Martin Neil, and Daniel Berger. "Bayes and the Law." *Annual Review of Statistics and Its Application* 3, no. 1 (2016): 51-77. doi:10.1146/annurev-statistics-041715-033428.

company owned 15% of the cabs. In addition, you were the witness to a hit-and-run accident where you claimed that the cab involved was green. You were then tested and it was found you could see the color of the cab correctly 80% of the time and incorrectly 20% of the time. Because of your ability to see the correct cab color 80% of the time, you assumed that the probability that the cab was green was 80%. However, as the name of this fallacy suggest, this ignores the base-rate, which in this case is the percentage of green cabs in the city. In order to arrive at the correct answer of 41%, we must take into account both the percentages of the blue cabs and the percentages of the green cabs and your ability to spot the correct color cab in the dark. Bayes' theorem allows us to do this. The Pr(Cab is green| you identified the color of the cab as green), or the probability that the cab is green given you identified the color as green, will give us the chances that the cab involved in the hit-and-run was actually green. When we apply the multiplication rule and Bayes' theorem, we can find that this is equal to the Pr(you identified the color of the cab as green given it was green)*Pr(the cab is green)/[Pr(you identified the color of the cab as green given it was green)*Pr(the cab is green)+ Pr(you identified the color of the cab as green given the cab was blue)*Pr(the cab is blue). In other words, you are finding the probability that the cab is green and you saw it as green divided by the overall probability you identified it as green whether it was actually green or not.

The percentages are explained as follows. The probability that you identified the color of the cab as green given the cab was Green is 80% since this means that you correctly identified the color. This is then multiplied by the overall probability that the cab is Green, which is 15%, since 15% of the cabs in the city are green. The product of these two, (80%*20%), is then divided by the overall probability that you identified the cab as green. This could occur if the cab was actually green or if you made a mistake and it was blue. Thus, you must calculate the probability you identified the color of the cab as green given it was Green and multiply it by the probability that the cab is green, which you just found was (80%*15%). Then you must add this to the product of the probability you identified the color of the cab as green given the cab was blue, or 20% since you made a mistake, and the probability of the cab being blue, or 85%, since 85% of the cabs in the city are blue.

If you place the percentages into the formula, then the probability that the cab is green given that you identified the color as green = (80%*15%)/[(80%*15%) + (20%*85%)]=41.37931….% or about 41%. These odds are quite a bit less than the 80% you assumed.[53]

Simpson's Paradox Solution

The mathematics behind Simpson's Paradox differs from the solutions that have previously been presented. However, the paradox is very prevalent in probability and statistics and incorporates several of the basic and conditional probability concepts that we have discussed.

Simpson's paradox occurs when two variables behave differently when considered separately than when they are considered jointly. In the Berkeley example, it appeared that the males were admitted at a higher rate than the females when the whole population was considered. However, when the admittance rates were observed for individual departments, the females appeared to have higher admittance rates in certain departments.

---

[53] The Cab scenario and the data behind it were adapted from the following source: Ibid.21.

In order to evaluate the data, researchers examined the six largest graduate departments, the number of applicants to each, and the percentage of male and female students accepted into them. The chart[54] below demonstrates these numbers. From this data, researchers were able to discover an extraneous variable that accounted for the paradox in the data. The female applicants disproportionately applied to the departments with lower overall admissions rates, while the male applicants disproportionately applied to the departments with higher admissions rates. In addition, it can be seen in the chart how in certain departments, female applicants had a higher admittance rate than the male applicants. This contradicts the idea that Berkeley was discriminating against gender and proves that trends as a whole can differ from the trends of their categorical parts. [55]

Table 1: Data From Six Largest Departments of 1973 Berkeley Discrimination Case

Department	Men		Women	
	Applicants	Admitted	Applicants	Admitted
A	825	62%	108	82%
B	560	63%	25	68%
C	325	37%	593	34%
D	417	33%	375	35%
E	191	28%	393	24%
F	272	6%	341	7%

Pearl describes this mathematically through a series of conditional probability inequalities.

---

[54] P.J. Bickel, E. A. Hammel, and J. W. O'connell. "Sex Bias in Graduate Admissions: Data from Berkeley." *Science* 187, no. 4175 (1975): 398-404. doi:10.1126/science.187.4175.398.

[55] Brad Hershbein, "When average isn’t good enough: Simpson’s paradox in education and earnings | Brookings Institution," Brookings, July 28, 2016, , accessed 2017, https://www.brookings.edu/blog/social-mobility-memos/2015/07/29/when-average-isnt-good-enough-simpsons-paradox-in-education-and-earnings/.

The inequalities are as follows:

Pr(E|C) > Pr(E| not C)

Pr(E|C, F) < Pr (E| not C, F)

Pr(E|C, not F) < Pr(E|not C, not F)

A paradox seems to occur if we think of this situation as a cause-effect scenario. For instance, let C represent taking a certain drug, E represent recovery, and F represent being female. The first inequality suggests that the drug is beneficial to the entire population. However, the bottom two inequalities suggest that the drug is harmful to both males, or not F's, and females. However, if we do not consider C and E to have a cause and effect relationship, then the paradox disappears. Instead, suppose that C represents an evidence for E, which Pearl states could be due to factors that cause both C and E. In other words, in our example, suppose that the drug appears beneficial to the whole population because males, who recover more quickly than females regardless of the drug, are also more likely than the females to take the drug. In this case, if a drug-taking patient is selected at random, then it is more likely that the patient is male and thus more likely to recover. This is consistent with all three of the inequalities and therefore solves the paradox by eliminating it. [56]

Expected Value Theory

Expected Value Theory Concept

Decision theory is centered on how an agent, typically an individual, reasons through their many different options and eventually makes a decision. This decision may

[56] Judea Pearl, "Understanding Simpson's Paradox," *SSRN Electronic Journal*, December 2013, doi:10.2139/ssrn.2343788.

be based on a number of different factors, including the agent's beliefs and desires. Among the different subsets of decision theory is normative decision theory, which is concerned with what criteria these beliefs and desires should satisfy in any generic circumstances and focuses on how to handle situations of uncertainty. The expected value theory is then a form of normative decision theory that explains that in scenarios where there is uncertainty, the agent should choose the option with the greatest expected desirability or value.[57]

The expected value has now come to be calculated by taking the weighted average of all possible outcomes in a given situation, with the weights being assigned by the likelihood, or probability, that any particular event will occur. In other words, in times of uncertainty, the expected value theory can be used to find which option may produce the highest possible value, or the amount of return or payoff.[58] While the theory may have its limitations, as in the case of risk, it may aide in the understanding and solving of certain paradoxes, such as the Allais paradox and St. Petersburg paradox.

Eighteenth century mathematicians Daniel Bernoulli and Gabriel Cramer first discussed the idea behind expected utility theory. They argued that the maximization of expected wealth was not enough to explain the decisions made by reasonable individuals faced with risky monetary options. After the consideration of examples similar to the Allais paradox, Bernoulli and Cramer suggested that these risky monetary options be evaluated not by their expected returns but by the expectations of the utilities of their returns. Utility in this case accounts for the amount of satisfaction or happiness derived from a particular

[57] Katie Steele and H. Orri Stefánsson, "Decision Theory," Stanford Encyclopedia of Philosophy, December 16, 2015, accessed 2017.
[58] Investopedia Staff, "Expected Utility," Investopedia, June 05, 2007, accessed 2017.

good or service. However, it is important to note that expected utility also does not fully account for how risk affects behavior. In addition, some have argued that in the long run, expected utility approaches expected value. [59] Let us consider the roles value and utility play in the following paradox solutions.

Allais Paradox Solution

The Allais paradox occurred when you entered into a casino and were asked twice to pick between two different gambles. The first gamble, gamble A, gave you a 100% chance of winning $100. Gamble B on the other hand gave you a 10% chance of winning $500, an 89% chance of winning $100, and a 1% chance of receiving nothing. You chose gamble A since it lacked risk. Next, you were given the choice between gamble C and gamble D. Gamble C gave you an 11% chance of receiving $100 and an 89% chance of receiving nothing. Gamble D gave you a 10% chance of receiving $500 and a 90% chance of receiving nothing.[60] You decided to go with gamble D.

As we will see using expected value theory, gamble B and gamble D have the highest expected payouts, meaning they appear to be what the reasonable person should choose as they would lead to the highest possible winnings. So why would a person choose A over B? Some may suggest that gamble A is better since it has no risk, while gamble B still gives you a small chance of receiving nothing. However, if a person were to have made the decisions based on risk, then they would have chosen gamble A and gamble C. Since they chose D in the second set of gambles, this contradicts this notion of only choosing based on risk. So

---

[59] Peter C. Fishburn, *The Foundations of Expected Utility* (Dordrecht, Holland: D. Reidel Pub. Co., 2010).

[60] Ibid. 25.

(Gamble A,B,C,D and the numbers behind them orginated from this source.)

why is it that we manage to choose the first gamble incorrectly, by expected value, and yet choose the second correctly?

Let us first evaluate why gamble B and gamble D are the decisions with the highest expected payouts. By expected value theory, each gamble is equal to the sum of the products of the probability of each outcome and the outcome. For gamble A, the only outcome is $100, which when multiplied by 100%, its probability of occurring, is equal to an expected payout of $100. Next, gamble B has an expected payout equal to (10%)*($500)+(89%)*($100)+(1%)*($0)=50+89=$139. Since $139 is greater than $100, the expected payout of gamble B is higher and is the gamble the rational person should choose. Following the first gamble, is the decision between gamble C and gamble D. Gamble C has an expected payout equal to (11%)*($100)+(89%)*($0)=$11. Gamble D has an expected payout equal to (10%)*($500)+(90%)*($0)=$50. Gamble D thus has a higher expected payout than gamble C since $50 is greater than $11. From these calculations, it is clear that gamble B and gamble D have higher likelihoods of greater winnings. After learning about expected value, you now have a tool to better evaluate which gambles may be best not only in a casino, but in other real world scenarios, such as the stock market.

As for those that still believe gamble A is the better bet as it lacks risk, then understand that the average person subconsciously places a higher utility on the absence of risk than it does on reduced risk; whether or not this is always rational, is up for debate.

St. Petersburg Paradox Solution

The St. Petersburg paradox is another paradox that employs the expected value theory. The paradox occurs when a person plays the St. Petersburg game, which involves flipping a coin until it lands on heads. If the number of flips that it takes to get a tail is equal to n, then n flips will you give $\$2^n$ worth of prizes. For instance, if you get a tail on the first flip, then you would win $\$2^1=\$2$. If you were to get a tail on the second flip, then you would win $\$2^2=\$4$ and so on. If you were to keep flipping heads, then the game could potentially continue seemingly indefinitely. This poses a problem because if the game could be played an almost infinite number of times, then the prizes could also grow to be arbitrarily large. Then, any entry fee into the game, no matter how high, would seem as though it would be worth paying, as it would always be less than an arbitrarily large sum of prize money. However, it seems irrational to assume that someone would be willing to pay an exceedingly high price for such a game. There must be a cutoff point, or a point at which the utility of the game is no longer increasing.[61]

Let us apply the expected value theory to see if we can find how many flips it would take to reach the highest expected payoff. This should reveal when we should stop playing the game. The expected value theory dictates that we multiply the probability of an event happening by the prize that would occur at that event. For the first event, there is a 1 in 2 chance of flipping a tails, since on a fair coin you can role a heads or a tails. The prize that you would win for flipping a tails on the first try would then be $2 as previously calculated. So, the expected payoff would be $(1/2)*(\$2)=\$1$. The second event, or flipping a tails on the second try, is equal to the probability of flipping a heads on the first try multiplied by the probability of flipping a tails on the second try, or $(1/2)*(1/2)=1/4$. The prize is $4, so the expected payout is $(1/4)*(\$4)=\$1$. If we continue this pattern, then we will find that the expected payout of every potential round in the game is $1. The expected value of the game is then the sum of each of these individual expected payouts. If they are each $1 and there

[61] Ibid. 28.

is the potential for a high number of rounds, then the sum becomes arbitrarily large. Thus, the expected value theory further supports the idea that playing the game, no matter the entry fee, could lead to an arbitrarily large prize. Therefore, it appears to be logical to keep playing. If you cannot accept the idea that there could be an exceedingly high entry fee worth paying due to the possibility of a larger and larger prize, then consider one of the offered solutions.[62]

While there have been many offered solutions to this problem, none are perfect. Consider however, as we did with the Allais paradox, the concept of risk, and by extension, utility. There is a point at which the rational person does not want to risk any more money on such a game. Ian Hacking in his 1980 paper believed this point to be at an entry fee of $25. Half of the time the game only pays $2 since there is a 50% chance of getting a tails on the first try. In addition, the probability of getting $4 or less can be found if we apply the complement rule and calculate 1-25% since there is a 25% chance the second flip is tails, which gives us 75%. As the odds of gaining each higher prize decrease rapidly, we can see how it seems riskier and riskier to invest more into the game. Hacking came up with a maximum entry fee of $25, since the odds of getting more than $25 is less than 1 in 25, or less than 4%. While this is arbitrary, it does represent one possibility for a finite answer. However, one issue with this is that some may be less risk averse than others.[63] In other words, there may be someone willing to buy into the game even if there is less than a 4% chance of winning. Regardless, you can either accept that the game can have an exceedingly high entry fee since the prize can be arbitrarily large, or you can weigh the expected value

---

[62] Ibid.

[63] Ian Hacking, "Strange Expectations," *Philosophy of Science* 47, no. 4 (1980): doi:10.1086/288956.

and expected utility of playing the next round and make a decision for yourself on when to stop playing.

Newcomb's Paradox Solution

You may recall that Newcomb's paradox involved the picking of either a box with $1,000 in it and/or a box with $1 million or nothing in it. The catch was that if a mystical being predicted you would choose both boxes, then it would put nothing in the second box. However, if it predicted you would just take the second box, then it would place $1 million in the box. The paradox may be set up with Bayes' Theorem in different ways, depending on which approach you take.

Let us first consider the players and the variables involved in Newcomb's paradox. You and the wise being, W represent the players. The variables are W's prediction, which we will refer to as g, and the choice you make, or y. There are two different versions of Bayes' theorem that may be interpreted from these variables.

The first approach has the equation $Pr(y,g)= Pr(g|y)*Pr(y)$. It would set up Bayes' theorem to base $Pr(y,g)$, the probability of the choice make and W's prediction, on the $Pr(g|y)$ and the $Pr(y)$, or the probability of W's prediction given the choice you make and the probability of the choice you make. This approach assumes that W's algorithm for predictions is never wrong and thus that W has the power to set the conditional probability. In other words, $Pr(g|y)$ represents the strategy W chooses and $Pr(y)$ represents the strategy that you choose. It is important to note that these strategies do not specify which choice was made, but the set of choices that can be made. After considering this, then the first approach ultimately tells you to make the decision by choosing the $Pr(y)$ that maximizes the expected payoff under the $Pr(y,g)$ associated with that choice. So, if the first approach assumes there is a 100% chance that W predicts correctly, then there is either an expected payoff with choosing both boxes of 100%(0+$1000)= $1000 or an expected payoff with choosing box B of 100%($ 1 million)=$1 million. Hence, the first approach's conclusion is to choose only box B.

The second approach has the equation Pr(y,g)= Pr(y|g)*Pr(g). It would set up Bayes' theorem to base Pr(y,g) on the Pr(y|g) and the Pr(g), or the probability of the choice you make given W's prediction and the probability of the choice W makes. The second approach assumes that your choice occurs after W has made its prediction and you do not know what that prediction is at the time of your choice. In other words, W cannot impact your prediction. This means that unlike in the first approach, W cannot affect the conditional. However, it does still have its own strategy.

Once again you must choose your strategy, so you consider the expected payoffs under the associated Pr(y,g). In this approach, choosing box B could result in an expected value equal to Pr(g=B)*($1 million) + Pr(g=AB)*($0) or the expected value that W guessed you would choose box B plus the expected value that W guessed you would choose box A and B. Thus, box B could payout once again either $1 million or $0, no matter the probabilities for W guessing you would only choose box B and W guessing you would choose both A and B. However, the expected value of choosing boxes A and B is equal to Pr(g=b)*($1,001,000) + Pr(g=AB)*($1,000). This is because if W was wrong and guessed you would choose only box B, then you would get $1,001,000 and if W guessed you would choose box A and B you would get $1,000. No matter the probabilities, this will have higher expected payout than just choosing box B since $1,001,000 is always going to be greater than $1 million and $1,000 is always going to be greater than $0. Thus, the second approach's conclusion is to choose boxes A and B. [64]

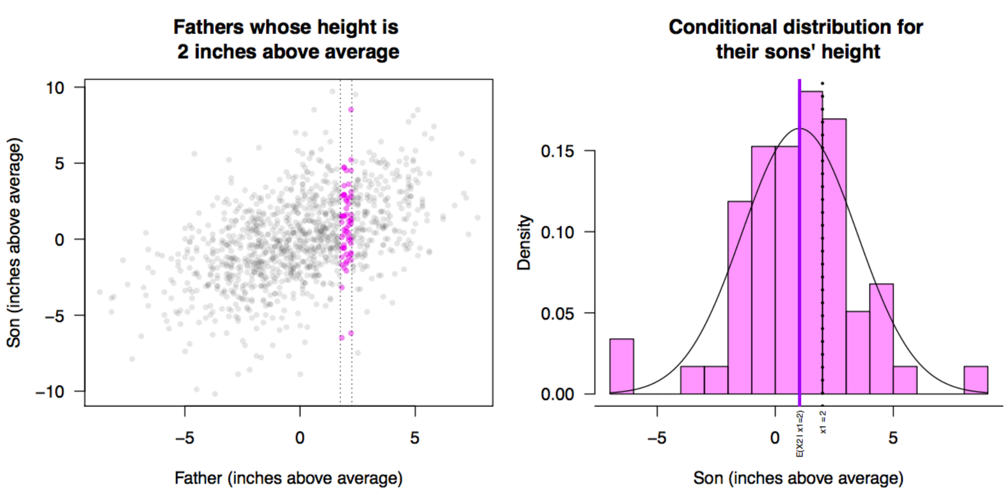
Whether you take the first approach or the second, it is key to understand why you want to choose the box or boxes you do and what could cause you to reconsider.

Regression To The Mean

Regression to the mean is based on the probability model known as the bivariate normal distribution. A graph of the normal distribution is shaped like a bell curve. The bivariate normal distribution is similar to this, but instead of graphing one variable, it graphs two correlated random variables, which we will call X_1 and X_2. This kind of graph is built on five types of parameters. The first two are the mean and variance of X_1 where the mean is the average and the variance is the measure of how widely individuals in a group vary. The second two are the mean and variance of the second variable, X_2, and the final parameter is the covariance of X_1 and X_2, which is the measure of how widely the two variables vary from each other, or rather the extent of their association. Suppose that we consider the heights of the fathers and sons that we discussed earlier and let X_1 be the fathers whose heights are about 2 inches above average. Then, we can use the bivariate normal distribution to find the average height that their sons would be. The charts that show this are demonstrated below. The first shows a plot where the dots in purple represent the father-son pairs that have a father 2 inches above average. These dots were then put into the histogram shown in the second chart. The histogram shows the conditional distribution, or the probability of the son's height given that the father is two inches taller than average. The normal distribution, or bell curve is then added to the

[64] Ibid.30.

histogram and appears to be a good fit as it shows that the conditional distribution is centered on 1. In other words, the sons with fathers two inches taller than the general population are on average, 1 inch taller than the general population. This shows the regression to the mean idea because the sons' heights are closer to the general population average than the fathers' heights are. This is because height involves heredity, which involves some luck. Outcomes, such as height or performance in a game, are impacted by a constant, or relatively constant component, such as ability in the case of professional athletes or genes in the case of fathers. However, such events are also impacted by a luck component. When outcomes are governed by both a constant and an element of luck, then the odds are that after one extreme outcome, such as a high scoring game or tall father, there will be an outcome closer to average.[65]



---

65 Ibid.38.

SECTION IV: REAL WORLD APPLICATIONS

Basic Probability: Applications of the Birthday Paradox

There are numerous applications of the birthday problem outside of determining the odds that at least two people in a room share the same birthday. It is particularly useful in situations where it is helpful to know the likelihood that two individuals will randomly share the same information.

For instance, you would not want your information to match or come close to matching that of a recent criminal's. Consider a case involving DNA evidence, such as the one we discussed with the prosecutor's fallacy. A crime was committed and the DNA from the scene was collected. Assume there is a 1 in 5 million chance that an innocent person would have DNA that matched the sample. Also, suppose there is a database with the genetic material of 10 million people. An officer enters in the newly collected evidence into the database to see if there is a match. While it is highly unlikely that you could pick a specific person that would happen to match the sample DNA, just as it would be rare to find someone with your exact birthday in a room of people, the odds are good that there exists some person in the large database that matches the sample. On average, the database will likely come up with two names that match the sample, with at least one of them being innocent. In this scenario, the larger the database, the higher the likelihood that more people will match, as is the case with the odds of people having the same birthday in a room with more people.

While it is discouraging to think that as DNA databases grow, more people will match with someone who committed a crime, there are laws that dictate the quantity of genetic loci that must match for the DNA evidence to be admitted in a trial. In addition, it is important not to discount the value of the database even if it does generate 10 possible matches for one given sample. This may narrow an investigation down to 10 people and officers may be able to narrow the search down further by determining which of the 10 could have been in the area at the time of the crime, which knew the victim, and/or which could have physically been able to commit the crime. That being said, it is imperative that judges and juries consider the countless other factors in a case aside from DNA evidence because if they do not, they run the danger of convicting an innocent person who happens to match the criminal in one aspect. [66]

Along the same lines, the birthday problem may also have applications in Class Phenotype Probability. Suppose you are given six characteristics, including blood type, RH positive/negative, sex, mid-digital hair positive/negative, earlobes attached/unattached, and PTC taste receptor. It is possible to determine the probability that a certain combination of these exists and the odds that two people share the same combination. This is immensely valuable in the medical field because finding phenotype matches can lead to finding matches between organ donors and potential recipients.[67]

In addition, the birthday paradox has many uses in computer science in the areas of cryptology and cyber security. An aptly named "birthday attack" occurs when a computationally intensive strategy is used to break encrypted signatures. During the attack, a "collision" occurs when different sets of data yield the same cryptographic hash value, where a hash value occurs when an arbitrarily large amount of data is mapped to a fixed amount of data. Throughout an attack, a hash-generating function is repeatedly

---

[66]Jordan Ellenberg, "If Police Find a DNA "Match," That Doesn't Mean They Have the Right Suspect," Slate Magazine, June 05, 2013, accessed 2017,

[67] Lidia Gonzalez, "Birthday Problem."

evaluated using random inputs until the output creates a collision with the true hash value it seeks to duplicate.

In other words, when we think of cryptography, we think of files being encrypted to protect private information. To break an encryption, an attack is initiated. The attack uses a tool called a hash-generating function. Hash-generating functions are very complicated, but let us suppose for the sake of example that the function is a sum, such as x+y. Large amounts of data can be plugged into x+y and different data will produce different results, such as 2+3=5, 4+8=12, etc. However, suppose x+y, when different data is inputted into it, keeps equaling 4. We would say that it collides at 4. Then, this would reveal that 4 is an important hash value that could be one of a series of important values that could lead to the correct set of data to decrypt a file.[68]

In any given situation where there is a chance that you or something you have could randomly match with someone else or something that they have, whether it is DNA, Social Security numbers, cyber security passwords and encryptions, medical characteristics, lottery numbers, etc., it is helpful to know how to determine the odds of it happening and understand that there is often a much greater chance of it happening than you may think, particularly in large samples.

Conditional Probability and Bayes' Theorem Applications

While Bayes' Theorem plays a key role in the solution to many of the paradoxes that were discussed, it is also applicable in numerous real world scenarios. However, it is often misinterpreted or not considered, which can be particularly detrimental in the field of medicine.

[68] Ibid.

One real world application where you could greatly benefit from a heightened knowledge of Bayes' theorem would be if you decided to go to the doctor's office and get a test to determine whether or not you were sick. Suppose that the test has a 99% chance of being accurate. In other words, 99% of people who test positive are sick and 99% of people who test negative are healthy. In addition, suppose it is also known that only 1% of people in the country are sick with this particular illness. If this is true, then what are the chances that you are sick if you test positive? The average person might answer that there is a 99% chance that you are sick since that is how reliable the test is. However, since only 1% of the population has the sickness, this cannot be true. The information given provides the probability that you test positive given that you are sick, but you need to know the probability that you are in fact sick given that you tested positive.

Bayes' Theorem states that in order to find the probability that you are sick given that you tested positive, you need to first find the probability of being sick and testing positive, and then divide this by the total probability of testing positive whether you are sick or healthy.

$$\Pr(\text{sick} \mid \text{test positive}) = \frac{\Pr(sickandtestpositive)}{\Pr(testpositive)}$$

It is helpful to break the question down into smaller Bayes' Theorem equations and then combine them. In order to find the probability that you are sick and you test positive, you would use Bayes' rule and the multiplication rule and find the product of the probability that you test positive given that you are sick, which is given to be 99%, and the probability that you are sick, which is given to be 1% since only 1% of the country is sick.

$\Pr(\text{sick and test positive}) = \Pr(\text{test positive} \mid \text{sick}) * \Pr(\text{sick}) = (99\%)(1\%)$

Then, you would divide this product by the overall probability that you test positive. The probability that you test positive on the test is equal to the probability that you are sick and test positive plus the probability that you are healthy and test positive, since you could be sick or healthy and still test positive.

Pr(test positive) = Pr(sick and test positive) + Pr(healthy and test positive)

The probability that you are sick and test positive you just found above, so you now need to find the probability that you are healthy and test positive. If you apply Bayes' Rule once again as well as the multiplication rule, then the probability that you are healthy and test positive is equal to the product of the probability that you test positive given that you are healthy, which is given to be 1%, since there is a 1% chance the test is unreliable, and the probability that you are healthy, which is given to be 99%, since there is only 1% of the country that is sick.

Pr(healthy and test positive) = Pr(test positive | healthy)*Pr(healthy)=(1%)(99%)

When the equations are combined, then the probability that you are sick given that you test positive is only 50%, meaning you really only have a 50% chance of being sick if you tested positive.[69]

$$\Pr(\text{sick} \mid \text{test positive}) = \frac{\Pr(\textit{sickandtestpositive})}{\Pr(\textit{testpositive})} = \frac{(0.99)(0.1)}{(0.99)(0.1)+(0.1)(0.99)} = 0.50 = 50\%$$

This is a significant difference from the 99% chance of being sick that you might have previously thought. From this scenario, you can see the importance of looking more closely at the accuracy of the medical tests you take and the prevalence of the disease in

[69] "What is Bayes's theorem, and how can it be used to assign probabilities to questions such as the existence of God? What scientific value does it have?," Scientific American: The Sciences, November 30, 2006, accessed 2017.

your area. While we would like to believe that medical professionals have the statistical knowledge to accurately interpret our health data, it is easy to see how mistakes can be made and how those mistakes can lead to false diagnoses.

In addition to medical applications, Bayes' theorem also has many uses in technology and data analytics. One task that it helps to accomplish that we could, and often do without realizing, benefit from daily is the filtering of our spam emails. Software that employs Bayes' theorem works by first asking the user to generate a database with words and symbols, such as the $ sign, certain IP addresses and domains, etc., that may be assembled from a sample of spam mail and legitimate mail, which is referred to as ham mail. The software will also likely include its own database of frequently occurring words and symbols. Once the user's personalized database has been created, then the software will assign probability values to each word or symbol based on how often these words and symbols occur in spam mail compared to ham mail. For instance, suppose the word "mortgage" appears in 400 out of 3,000 spam emails and 5 out of 300 ham emails. The software would then assign a probability value to the word by calculating the Bayes' Theorem. The probability that an email is spam given that it contains the word "mortgage" is equal to the probability that an email is spam and has the word "mortgage" in it divided by the probability that an email contains the word "mortgage." This is equal to $\dfrac{\frac{400}{3000}}{\frac{400}{3000}+\frac{5}{300}}$, which is approximately 0.8889. Once the probability values of many words and symbols have been calculated, software can evaluate a new incoming email. First, it will compile the probability values of the key words in the email, and then it will assign the email with an overall probability of being spam. If this probability is over a certain amount, such as 0.9, then it is designated as spam mail and is filtered. While there are some errors to the system, the Bayesian approach tends to have a high success rate. In addition, it is favored due to its ability to take an entire message into account and because of its way of self-adapting by adding more key words and symbols to its database.[70]

Whether Bayes' theorem guides you in the important task of interpreting your medical results or saves you a bit of time sifting through your inbox, it is a valuable mathematical concept with vast real-world implications.

Expected Value Theory Applications

Perhaps the most common application of expected value theory, which in some cases is controversial, occurs in the stock market. While it may be impossible to perfectly predict how stocks and the market itself will behave, expected value theory provides a way to quantitatively value one stock or portfolio over another. These values assigned to stocks are referred to as expected returns. The expected return is the amount of profit or loss an investor anticipates on an investment that has various known or expected rates of return. In other words, it is nothing more than the investment's expected value. The reason this is controversial is because when you consider the expected value formula, you must have the quantitative value of the outcomes and the probability of those outcomes occurring. In the stock market, both the outcomes and the probabilities can be unpredictable, since no one knows for sure, except in certain cases, what the dividends, or outcomes will be, or what the chances of those particular dividends being produced are. Much of this information is

[70] "Why Bayesian Filtering Is the Most Effective Anti-spam Technology," 2011, Rep. GFI Software, accessed 2017.

based on historical data, which is not always guaranteed to be the same or even similar to what today's data will be. Another reason this method remains controversial is that it does not take into account the full extent of risk. While it weighs probabilities, if an expected return is high, but has a very small chance of occurring, then it may not be the safest investment. Even despite these shortcomings, expected returns are frequently quoted and compared among investors considering buying and selling different stocks.[71]

In addition to helping value stocks, expected returns are also helpful in valuing portfolios when the expected returns of the stocks comprising the portfolio are known. For instance, consider you have a portfolio composed of stocks A, B, and C. Suppose you invested $500,000 in stock A with an expected return of 15%, $200,000 in stock B with an expected return of 6%, and $300,000 in stock C with an expected return of 9%. Then just as expected value is calculated, the outcomes are weighed by their probabilities and then summed. The only difference is that the investment amounts are converted to percentages so that the total portfolio's expected return is a percentage. In other words, $500,000 is equal to 50% of $1,000,000, or the total value invested. $200,000 is then equal to 20% and $300,000 is equal to 30%. Thus, the equation becomes $(50\%*15\%)+(20\%*6\%)+(30\%*9\%)=7.5\%+1.2\%+2.7\%=11.4\%$. [72]This percentage can then be used to determine how this portfolio stacks up against others and how its returns compare to an investor's desired returns. As long as the assumptions made about the returns as well as the component of risk are considered, the expected returns may be

[71] Investopedia Staff, "Expected Return," Investopedia, April 17, 2015, accessed 2017.
[72] The data from this example comes from the following source: Ibid.

incredibly valuable to the wise investor. This is why it is important to know what goes into calculating them and the expected value theory behind doing so.

Since we discussed expected utility in addition to expected value, let us consider an expected utility application. One such application of the expected utility theory is in the study of climate change. Matthew Kahn and Daxuan Zhao published a paper in February of 2017 on how the prevalence of so-called "climate skeptics" has led to reduced demand and support for products, services, and legislation that aide in climate adaptation. With a lack of demand from consumers for climate-resilient products, companies are less likely to focus on dedicating their resources to solving challenges in this field. Both professors argue that the "market potential" for climate adaptation is key because capitalism could incentivize necessary climate change innovations. These innovations they claim need to come in the way of more efficient air conditioning to protect us from climate change-induced heat, as well as in the way of architecture to reduce the flood damage to real estate. Their model on this is based on expected utility.

They claim that each person in the economy wants to maximize their expected utility, which in the model was calculated by multiplying the value of a person's future consumption by their chance of survival. In other words, they used expected utility to find the value of a person enjoying their life multiplied by the chance that they will live to enjoy it. This is connected to climate change because products aimed at climate change adaptation may impact an individual's survival rate. The model is designed to show that a rational person would demand these types of products in an effort to increase their chances of surviving and enjoying future consumption. However, around 50% of people in the United States are skeptical of human activity's contribution to climate change. These skeptics also aim to maximize their future utility, but do not believe that climate-adaptation products will alter their chances of survival and as a result, have no reason to buy such products. The model demonstrates that more skeptics result in less demand and a smaller available market for these products. A smaller available market then results in fewer innovations in climate adaptation and fewer entrepreneurs willing to dedicate their careers such innovations.[73]

Expected value theory and expected utility theory, along with being useful in the financial and scientific realms, may be useful in considering any decision with a series of choices. The math behind both may provide a way to more clearly define what the choices are and how they compare to one another. Even something as simple as a decision on whether to take the bus could be more thoughtfully considered through the use of both theories. While neither includes everything that should be considered in a decision, such as risk, the appropriate calculation of expected value or expected utility could lead you to make more prudent decisions.

Regression To The Mean Applications

As we have discussed, the regression to the mean phenomenon occurs when an extreme outcome is followed by a more average outcome and where the outcomes are governed, at least in part, by chance. There are many different real world applications for this. We have considered its occurrence in heredity, which is how it was first discovered, but we also discussed its relevance in a sports setting. Major league baseball players that have the highest batting averages in the first season are more likely to bat closer to the league's average in their second season, which many mistake for a slump. Along with

[73] Matthew E. Kahn and Daxuan Zhao, "The More Climate Skeptics There Are, the Fewer Climate Entrepreneurs," Harvard Business Review, March 16, 2017, accessed 2017.

heredity, the sophomore slump and the *Sports Illustrated* curse, regression to the mean also appears in the medical field.

Consider the many different measurements that a Doctor takes during a routine check-up. They are likely, at minimum, to take down your weight, cholesterol, and blood pressure. If one of these is in one of the extremes, either too high or too low, then it may be an indicator of an underlying disease or be a risk factor of a disease. For this reason, people with high blood pressure for example, will be treated with medication designed to lower the value closer to the average. However, if these same people come in and get measured again, then it is likely that their values will be closer to the general population's average. This, however, does not indicate that the medication they were given was effective. Even if these people remained untreated, it is more than likely that their blood pressure would fall closer to the general population's average because of regression to the mean. Therefore, we must be careful considering a medication's true effectiveness and make sure to account for extreme values naturally becoming closer to average values. [74,2]

The regression to the mean may also come up in publication bias. Rousseeuw explains that referees who determine which papers are submitted for publication do not always agree which papers should be accepted.[75] Because this then means that referees' judgments are made with error, then the quality of the paper cannot be perfectly correlated with the referees' judgments. Therefore, when the referees pass the chosen papers, which are supposedly the best, to the editor for publication, then the average quality of the paper

---

[74] J. M. Bland and D. G. Altman, "Statistics Notes: Some examples of regression towards the mean," *Bmj* 309, no. 6957 (1994): , doi:10.1136/bmj.309.6957.780.

[2] C. E. Davis, "The Effect Of Regression To The Mean In Epidemiologic And Clinical Studies," *American Journal of Epidemiology* 104, no. 5 (1976): doi:10.1093/oxfordjournals.aje.a112321.

[3] PJ Rousseeuw , "Why the wrong papers get published," *Chance*, 1991.

will be less than the editor thinks. In addition, the average quality of the rejected papers will be higher than the editor thinks. Thus, if you find that one of your papers was rejected, then you may not need to feel too upset. You could have just been a victim to the regression of the mean.

Regression to the mean can also occur in our interactions with other people. Consider the story of the small private college in the US that conducted an elaborate nationwide search for a new dean. The college first considered its internal candidates, but none of them were seen as a good fit. The search committee then sifted through hundreds of external applicants' résumés and references and identified several candidates that they felt were worth an interview. The committee then interviewed these candidates at neutral airports before settling on three to bring to campus. Once on campus, the three candidates were each subjected to two days of meetings with the faculty, administration, and students. While the search committee was initially enthusiastic about their final three, the candidates each visited the campus with limited success. The committee and those who had participated in the interview process were instead disappointed. This occurred because although the three candidates appeared to be the best, they were almost surely not as good as they seemed to be. They were likely closer to average. This also explains why internal candidates are often at a disadvantage when upper-level positions are vacant. Internal candidates are a familiar commodity and so their abilities and their flaws are already known unlike their external counterparts.

Unfortunately, Smith says this same idea is true in our search for soul mates. While everyone looks for different things, let us assume that "pizzazz" encompasses them all. Of the many dozens, or hundreds, or thousands of people that we may meet, only a few will stand out. From these few, we may recognize signs of the so-called “pizzazz.” After doing so, we may then gather the courage to get to know a portion of the few. However, when we do, Smith says that we will likely be disappointed. They will not be as good as we thought they were, but more average. Smith offers that the bad news is that this is what we should come to expect, but that the good news is that we can keep looking. I prefer to consider the opposite, perhaps those we dismiss firsthand as not having the “pizzazz” may be better than we think. Smith also makes the point that the other person probably feels the same way about us.[76]

The regression to the mean may also be an issue that impacts safety studies. Speeding is something many of us may be guilty of, but how much does a traffic camera impact our behavior? Several studies have been completed to show the effectiveness of engineering equipment on reducing speeding individuals. However, Park and Lord suggest that it is not necessarily the speed equipment that has led to fewer violations. It may instead be that extreme outcomes, which were not normal for an area, were occurring before the equipment was installed. In other words, people were speeding in a certain area more than they normally would. This ceased once the equipment was put into place, so those studying it believed that the equipment caused the speeding to decrease. However, it is likely that speeding was occurring at higher than normal rates for the area before the equipment was installed, which skewed the results. Park and Lord concluded that in order

[76] Gary Smith, "What does regression to the mean, mean?," ABC - Australian Broadcasting Corporation, May 12, 2015, accessed 2017,

to test the effectiveness of such equipment, more rigorous statistical methods that would take into account regression to the mean should be used.[77]

[77] Peter Park and Dominique Lord, "Investigating Regression to the Mean in Before-and-After Speed Data Analysis," *Transportation Research Record: Journal of the Transportation Research Board* 2165 (2010): , doi:10.3141/2165-06.

SECTION V: CONCLUSION

Paradoxes and fallacies represent occasions where our logic has failed us. Whether we are trapped between two conflicting answers or lost in an error, something has gone wrong. However, this failure is often not a bad thing. The book, *The 5 Elements of Effective Thinking,* even claims seemingly paradoxically that failing is one of the five keys to our success. In order to fail productively, we must first try something, then recognize and assess our errors, and then make an improved attempt. Eventually, a steady stream of improved attempts can lead to a discovery or an innovation, as in the case of Thomas Edison and the invention of the light bulb.[78]

On a small scale, this thesis serves as an example of this process. In one sense, the writing of this thesis represents productive failure. Numerous drafts were written, and after each, more errors in content and grammar were recognized and corrected. From these corrections, came improved attempts. Ideally, this attempt is a success because it represents the most improved of the attempts and is one step closer to reaching its goal of bettering its readers' decisions.

In addition, the content of this thesis goes through the steps of productive failure. In the beginning, through the discussion of paradoxes and fallacies, this thesis seeks to explain common failures in logic, and why and when they occur in the hopes that readers may more easily recognize them in the future. Then, the thesis moves on to discuss how these errors might be resolved through the use of a few key mathematical tools. These tools were

---

[78] Edward B. Burger and Michael Starbird, *The 5 elements of effective thinking* (Princeton, NJ: Princeton University Press, 2012).

designed to give the reader the ability to make improved attempts when they are confronted with paradoxes and fallacies in their own lives. Steadily, these improved attempts could lead to more productive failures in decision-making, and hopefully, to the innovation of a new thought or new solution.

Among those likely to be faced with many new incoming obstacles and opportunities to fail, are recent graduates entering the labor force. By incorporating paradoxes and fallacies and the mathematics behind them into the degrees of mathematics and hard sciences students, as well as liberal arts and humanities students, colleges and universities could provide all of their graduates with a valuable education that combines the knowledge of a certain field or fields, with the ability to tackle failure. With this power, graduates will be more prepared to fail their way to success and innovation in the real world. If what starts here changes the world, let us consider making this addition to our curriculum at the University of Texas.

“That’s All Folks!”

SECTION VI: BIBLIOGRAPHY

"Allais Paradox," Policonomics, 2012. Accessed 2017.

Ball, W. W. Rouse. *Mathematical Recreations and Essays*. London: Macmillan, 1959.

Bar-Hillel, Maya. "The Base-Rate Fallacy in Probability Judgments." *Acta Psychologica* 44, no. 3 (1980): 211-33. doi:10.1016/0001-6918(80)90046-3.

Barnett, A. G. "Regression to the mean: what it is and how to deal with it." *International Journal of Epidemiology* 34, no. 1 (2004): 215-20. doi:10.1093/ije/dyh299.

Benford, Gregory, and David Wolpert. "What Does Newcomb's Paradox Teach Us?" *SSRN Electronic Journal*, 2010. doi:10.2139/ssrn.1381295.

Bickel, P. J., E. A. Hammel, and J. W. O'connell. "Sex Bias in Graduate Admissions: Data from Berkeley." *Science* 187, no. 4175 (1975): 398-404. doi:10.1126/science.187.4175.398.

Bland, J. M., and D. G. Altman. "Statistics Notes: Some Examples of Regression Towards the Mean." *Bmj* 309, no. 6957 (1994): 780. doi:10.1136/bmj.309.6957.780.

Briggs, Rachael. "Normative Theories of Rational Choice: Expected Utility." Stanford Encyclopedia of Philosophy. August 08, 2014. Accessed 2017.

"Bugs Bunny - Tortoise Beats Hare," Cartoonsonnet. Accessed 2017.

Burger, Edward B., and Michael P. Starbird. *The 5 Elements of Effective Thinking*. Princeton, NJ: Princeton University Press, 2012.

Burger, Edward B., and Michael P. Starbird. *The Heart of Mathematics: An Invitation to Effective Thinking*. Hoboken, NJ: Wiley, 2013.

Carlson, Bruce W. "Simpson's paradox." Encyclopædia Britannica. August 31, 2016. Accessed 2017.

Cohen, Marc S. "Zeno's Paradox of the Arrow." Zeno's Paradox of the Arrow. 2003. Accessed 2017.

"Conditional Probability, The Monty Hall Problem," Conditional Probability, 2008, accessed 2017.

Crockett, Zachary. "The Time Everyone “Corrected” the World’s Smartest Woman." Priceonomics. August 2016. Accessed 2017.

Davis, C. E. "The Effect Of Regression To The Mean In Epidemiologic And Clinical Studies." *American Journal of Epidemiology* 104, no. 5 (1976): 493-98. doi:10.1093/oxfordjournals.aje.a112321.

Ellenberg, Jordan. "If Police Find a DNA “Match,” That Doesn’t Mean They Have the Right Suspect." Slate Magazine. June 05, 2013. Accessed 2017.

Fenton, Norman, Martin Neil, and Daniel Berger. "Bayes and the Law." *Annual Review of Statistics and Its Application* 3, no. 1 (2016): 51-77. doi:10.1146/annurev-statistics-041715-033428.

Fishburn, Peter C. *The Foundations of Expected Utility*. Dordrecht, Holland: D. Reidel Pub. Co. 2010.

Gardner, Martin. "Free Will Revisited, With a Mind-Bending Prediction Paradox by William Newcomb." *Scientific American*, July 1973.

Gardner, Martin. *Hexaflexagons and Other Mathematical Diversions: The First Scientific American Book of Puzzles & Games*. Chicago, IL: University of Chicago Press, 1989.

Gardner, Martin. "Paradoxes Dealing with Birthdays, Playing Cards, Coins, Crows, and Red-haired Typists." *Scientific American*, April 1957.

Gardner, Martin. "Reflections on Newcomb's Problem: a Prediction and Free-Will Dilemma." *Scientific American*, March 1974.

"GFI White Paper Why Bayesian filtering is the most Effective Anti-spam Technology," 2011. Accessed 2017.

Gonzalez, Lidia. "Birthday Problem." 113-14.

Hacking, Ian. "Strange Expectations." *Philosophy of Science* 47, no. 4 (1980): 562-67. doi:10.1086/288956.

Hershbein, Brad. "When average isn't good enough: Simpson's paradox in education and earnings | Brookings Institution." Brookings. July 28, 2016. Accessed 2017.

Kahn, Matthew E., and Daxuan Zhao. "The More Climate Skeptics There Are, the Fewer Climate Entrepreneurs." Harvard Business Review. March 16, 2017. Accessed 2017.

Kahn, Sal. "The Monty Hall Problem." Khan Academy. Accessed 2017.

Kahneman, Daniel and A. Tversky, "On prediction and judgment." *Oregon Research Institute Bulletin*, 1972.

Kaplan, Daniel. *Statistical Modeling: A Fresh Approach*. Lexington: S.n., 2013.

Kramer, Edna E. *The Nature and Growth of Modern Mathematics*. Greenwich, CT: Fawcett Publications, 1970.

Lane, David M. "Regression Toward the Mean." OnlineStatBook. 2016. Accessed 2017.

Martin, Robert. "The St. Petersburg Paradox." Stanford Encyclopedia of Philosophy. November 04, 1998. Accessed 2017.

"Maurice Allais." Policonomics. 2012. Accessed 2017.

Meehl, Paul E., and Albert Rosen. "Antecedent probability and the efficiency of psychometric signs, patterns, or cutting scores." *Psychological Bulletin* 52, no. 3 (1955): 194-216. doi:10.1037/h0048070.

Nodjimbadem, Katie. "What Gives Bugs Bunny His Lasting Power?" Smithsonian.com. July 27, 2015. Accessed 2017.

Nozick, Robert. "Newcomb's Problem and Two Principles of Choice." *Essays in Honor of Carl G. Hempel*,1969, 114-46. doi:10.1007/978-94-017-1466-2_7.

Parco, Nicholas. "A look at victims of the Sports Illustrated cover jinx." NY Daily News. March 23, 2016. Accessed 2017.

Park, Peter, and Dominique Lord. "Investigating Regression to the Mean in Before-and-After Speed Data Analysis." *Transportation Research Record: Journal of the Transportation Research Board* 2165 (2010): 52-58. doi:10.3141/2165-06.

Pearl, Judea. "Understanding Simpson's Paradox." *SSRN Electronic Journal*, December 2013. doi:10.2139/ssrn.2343788.

Rousseeuw, PJ. "Why the wrong papers get published." *Chance*, 1991, 41-43.

Savant, Marilyn vos. "Game Show Problem." Marilyn vos Savant. Accessed 2017.

Smith, Gary. "What does regression to the mean, mean?" ABC - Australian Broadcasting Corporation. May 12, 2015. Accessed 2017.

Staff, Investopedia. "Expected Return." Investopedia. April 17, 2015. Accessed 2017.

Staff, Investopedia. "Expected Utility." Investopedia. June 05, 2007. Accessed 2017.

Stark, Philip B. "Probability: Axioms and Fundaments." Statistics at UC Berkeley. Accessed 2017.

Steele, Katie, and H. Orri Stefánsson. "Decision Theory." Stanford Encyclopedia of Philosophy. December 16, 2015. Accessed 2017.

The Editors of Encyclopædia Britannica. "Richard von Mises." Encyclopædia Britannica. September 27, 2002. Accessed 2017.

"The Tonight Show with Johnny Carson." CornellCast. Accessed 2017.

Thomas, Rachel. "Mathematical Mysteries: Zeno's Paradoxes." Mathematical Mysteries: Zeno's Paradoxes. December 1, 2000. Accessed 2017.

Thompson, William C., and Edward L. Schumann. "Interpretation of Statistical Evidence in Criminal Trials: The Prosecutor's Fallacy and The Defense Attorney's Fallacy." *Law and Human Behavior* 11, no. 3 (1987): 167-87. doi:10.1007/bf01044641.

Tierney, John. "Behind Monty Hall's Doors: Puzzle, Debate and Answer?" The New York Times. July 20, 1991. Accessed 2017.

Weisstein, Eric W. "Monty Hall Problem." WolframMathWorld. Accessed 2017.

"What is Bayes's theorem, and how can it be used to assign probabilities to questions such as the existence of God? What scientific value does it have?" Scientific American: The Sciences. November 30, 2006. Accessed 2017.

SECTION VII: BIOGRAPHY

Byrn Rathgeber was born in 1995 in Austin, Texas. She graduated from Westlake High School, and then enrolled in The University of Texas at Austin in 2013. Once on the forty acres, she chose to major in Plan II and Mathematics with a certificate in Business Foundations. After graduation, she has accepted an offer to work for Gerson Lehrman Group. She later intends to pursue a Masters in Business.